Full Report of the Comparisons of Administrative Record Rosters to Census Self-Responses and NRFU Household Member Responses

by

Mary H. Mulry
U.S. Census Bureau

Cristina J. Tello-Trillo
U.S. Census Bureau

Thomas Mule
U.S. Census Bureau

Andrew Keller
U.S. Census Bureau

Abstract

One of the U.S. Census Bureau's innovations in the 2020 U.S. Census was the use of administrative records (AR) to create household rosters for enumerating some addresses when a self-response was not available but high-quality ARs were. The goal was to reduce the cost of fieldwork during the Nonresponse Followup operation (NRFU). The original plan had NRFU beginning in mid-May and continuing through late July 2020. However, the COVID-19 pandemic forced the delay of NRFU and caused the Internal Revenue Service to postpone the income tax filing deadline, resulting in an interruption in the delivery of ARs to the U.S. Census Bureau. The delays were not anticipated when U.S. Census Bureau staff conducted the research on AR enumeration with the 2010 Census data in preparation for the 2020 Census or during the fine tuning of plans for using ARs during the 2018 End-to-End Census Test. These circumstances raised questions about whether the quality of the AR household rosters was high enough for use in enumeration. To aid in investigating the concern about the quality of the AR rosters, our analyses compared AR rosters to self-response rosters and NRFU household member responses at addresses where both ARs and a self-response were available.

Keyword*: 2020 U.S. Census, Nonresponse Followup

Abstract

One of the U.S. Census Bureau's innovations in the 2020 U.S. Census was the use of administrative records (AR) to create household rosters for enumerating some addresses when a self-response was not available but high-quality ARs were. The goal was to reduce the cost of fieldwork during the Nonresponse Followup operation (NRFU). The original plan had NRFU beginning in mid-May and continuing through late July 2020. However, the COVID-19 pandemic forced the delay of NRFU and caused the Internal Revenue Service to postpone the income tax filing deadline, resulting in an interruption in the delivery of ARs to the U.S. Census Bureau. The delays were not anticipated when U.S. Census Bureau staff conducted the research on AR enumeration with the 2010 Census data in preparation for the 2020 Census or during the fine tuning of plans for using ARs during the 2018 End-to-End Census Test. These circumstances raised questions about whether the quality of the AR household rosters was high enough for use in enumeration. To aid in investigating the concern about the quality of the AR rosters, our analyses compared AR rosters to self-response rosters and NRFU household member responses at addresses where both ARs and a self-response were available.

1. Introduction

The U.S. Census Bureau's approach to the 2020 Census included the use of administrative records (ARs) to reduce the cost of Nonresponse Followup (NRFU), which is the field operation where enumerators visit addresses that did not self-respond and attempt to obtain an interview. The 2020 Census is the first census to use ARs to enumerate some households. The use of ARs has become feasible because of advancements in computer capacity and processing speed.

One way that ARs were used in the 2020 Census was to enumerate an address for which a response was not by submitted by internet, mail, or telephone during the self-response period and no response was obtained in one visit by an enumerator during NRFU. The ARs for the address were required to meet a specified quality standard. When these quality criteria were met, no further visits were made to the address during NRFU.

The COVID-19 pandemic caused delays in the data collection and processing for the 2020 Census. The original plan for the Census called for the collection of self-responses to occur from March 12 to July 31, 2020, but the actual self-response period occurred from March 12 to October 15. NRFU was originally scheduled from May through July but was conducted with a soft launch in July through October 15.

In addition, the pandemic caused federal agencies to postpone the collection and delivery of some of the ARs that they had agreed to provide. These delays resulted in the Census Bureau staff postponing the start of the modeling that supported the AR enumeration from early May to late May. The agencies made monthly deliveries that contained updates of their Ars, which allowed the Census Bureau's AR processing staff to see if these files changed from month to month, and when a change occurred, to see the effect on the AR rosters.

The combination of the delay of NRFU data collection and the delay of the receipt of ARs created situations that did not appear when Census Bureau staff were conducting the research with the 2010 Census data in preparation for the 2020 Census. These circumstances raised questions and created opportunities for improving the census count using ARs beyond the original plans. In fact, the Census Scientific Advisory Committee recommended that the Census Bureau conduct analyses that compared census rosters and administrative records rosters for addresses where both types of rosters were available (Census Scientific Advisory Committee 2020).

This document focuses on the addresses where both a census roster and an AR roster were available, but the two rosters differed on the size of the household. Of particular interest is the situation where the census roster lists one more or one less person than the ARs identify as residing at the address since these were the types of discrepancies between the census roster and the AR roster that occurred most often. When an address had both an AR roster and a census self-response or a NRFU household member response, the response submitted by the household was the one that was used for the census enumeration in most circumstances. The topics covered in the document are as follows:

- Section 1 is the introduction to the topic.
- Section 2 has background information on the methodology and implementation of AR enumeration.
- Section 3 includes a description of the methodological approach used in the study.
- Section 4 contains results of analyses of the addresses where the census and AR rosters differ on household size.
- Section 5 discusses plausible explanations for why the census roster and the AR roster for an address differ by one person for some addresses.
- Section 6 contains a summary.

Understanding the characteristics of households where the census roster and AR roster have a different household size will add to the knowledge base for the use ARs in the most effective ways in future censuses. Of particular interest are the characteristics of the individuals who are on one roster but not the other when the size of the rosters differs by one. Our research seeks inform planning for the 2030 U.S. Census through these analyses.

2. Background

The Census Bureau has a long history of seeking ways of improving census-taking methodology and reducing cost while preserving data quality. For example, early in the 1940s, a clerical match between the 1940 Census and draft records found more males 21 to 35 years old registered for the draft than counted in the 1940 Census. The estimates were 14.9 percent more Black males and 2.8 percent more non-Black males (Price 1947). This result led to the recognition of the need to evaluate census coverage and prompted the development of the census coverage evaluation methodologies Demographic Analysis (Coale 1955) and Post-Enumeration Survey (Marks and Waksberg 1966). Both methodologies have evolved over the years, and both are part of the program to evaluate the coverage of the 2020 Census.

The Census Bureau has increased the use of ARs in its programs over the years as the quality of ARs and the computer technology to process large datasets has improved. In addition, sharing of AR data between federal agencies has been encouraged by the Office of Management and Budget, and thereby become more frequent under strict guidelines for protecting privacy. The source agency also must approve the way the receiving agency plans to use the data. Recent U.S. censuses have used ARs in some aspects of census taking but not for enumeration of households. The challenge in using ARs for census enumeration has been that the U.S. does not have a single AR system that covers the entire population.

As the planning for the 2020 Census started after the evaluation of the 2010 Census, the Census Bureau searched for ways to reduce the cost of Nonresponse Followup (NRFU) while maintaining the quality of the enumeration. NRFU is the field operation where enumerators attempt to obtain census responses from households that did not submit a self-response by internet, mail, or telephone. NRFU is one of the most expensive operations in the census. ARs appeared to be a potential source of data that could be used for census enumeration. Therefore, the Census Bureau conducted a research program from 2012 to 2018 to identify a cost-effective way to use ARs to reduce the cost of NRFU while preserving quality and protecting personally identifiable information from disclosure.

The research program produced a methodology for creating an AR roster for the household at an address in NRFU and assessing its quality. The quality assessment relied on three models to assign a status of AR Vacant, AR Nonresidential, AR Occupied, or No Determination to addresses in NRFU. AR rosters were created for addresses with the status of AR Occupied and assigned a quality status. An AR roster that met the quality standard was used for enumeration only if a self-response was not received for the address and one NRFU contact attempt was unable to resolve the status of the address. In addition, the AR roster had to meet a specified quality standard. For more details, see Section 2.3.3.

2.1 Population Under Study

According to census residency rules, the correct location for a person's enumeration is the address of the person's usual residence around Census Day, which is April 1 of the census year. Our population under study is defined as addresses where the Census Day residents were enumerated by self-response or NRFU household member response when an AR roster also was available. Another requirement for AR enumeration and therefore inclusion in this study is that the address appears on the Census Bureau's Master Address File (MAF) with an assigned identification number called a MAFID. The assignment of a MAFID indicates that the address has been verified. In addition, the AR roster must have one of the six valid household compositions (between one and three adults, with or without children) to be considered for AR enumeration (Mule, Keller, Konicki 2018). Previous research has shown that AR rosters for households containing more than six people tend not to meet accuracy requirements (Mule, Keller, Konicki 2018).

2.2 Matching Administrative Records using PIKs

The linking of the administrative records data to other data sources, which includes the self-response data, requires that both sources include Protected Identification Keys (PIKs). These

PIKs are essentially encrypted Social Security numbers or Individual Tax Identification numbers, which are included when using the term Social Security numbers in this paper. Administrative records data comes with Social Security numbers that the Census Bureau staff convert to PIKs after a validation of their accuracy through matching to Social Security Administration files, a procedure called the Person Identification Validation System (PVS) (Wagner and Layne 2014). When a data file with records for people does not come with Social Security numbers, the Census Bureau uses its system to match each person's information to reference files to assign the person's Social Security number and then encrypt it by assigning a PIK.

Sometimes the PVS fails to assign a PIK to a record. For example, 90.3 percent of the 2010 Census enumerations received a PIK from the PVS, but only 97 percent of the enumerations had enough information for an attempt to assign a PIK (Wagner and Layne 2014). Evaluation studies have shown that missing date of birth in a record is highly correlated with the PVS not assigning a PIK. In addition, an incomplete or fake name in a record is highly correlated with a PIK not being assigned (Wagner and Layne 2014; Mulrow et al. 2011).

2.3 Data

2.3.1 Data sources for AR rosters

The following AR sources were used to construct the AR rosters in a process that included identifying multiple records for the same person at an address and removing the duplicate records:
    (1) IRS 1040 forms filed for 2019
    (2) IRS 1099 forms filed for 2019
    (3) Medicare records
    (4) Indian Health Service Patient Database
    (5) Household Composition Key File at the U.S. Census Bureau.

All the sources listed above are self-explanatory except Household Composition Key File (Deaver 2020), which is a database created and maintained by Census Bureau staff using applications for Social Security numbers (SSNs) from the Social Security Numerical Identification (Numident) File. The Census Bureau maintains a Census Numident file derived from the Social Security Numident File that contains names and SSNs that are used in assigning PIKs to records to enable child-to-parent linking. A record for a child in the Household Composition Key File includes the names of the child's mother and father, if available, but not the SSNs of the parents.

During the construction of the AR rosters for census enumeration, Census Bureau staff attempted to assign a mother PIK and father PIK using the names of the parents. The Census Numident file was used to obtain age and sex information for each person in the administrative record household roster. Therefore, while each child record received a PIK based on the child's SSN, not all parent records were assigned a PIK. In addition, there may have been errors in a parent's PIK assignment (Mule, Keller, Konicki 2018).

An issue with the Household Composition Key File is that it does not provide address information; the file only provides the child's PIK and the PIKs of the parents if they could be assigned. To place children from the Household Composition Key File at addresses, the child was added to the AR roster for each of the addresses where either the mother or the father appears. If the mother and father appear in ARs at separate addresses, then staff added the child to each of these addresses. Similarly, if a parent appeared in multiple addresses in AR, then staff added the child to each of those addresses (Mule, Keller, Konicki 2018). One reason for adding the child to all the AR rosters listing one or both parents is that staff did not know which of the AR rosters, if any, would be used for enumeration.

A child being added to the AR roster at each parent's address does not necessarily produce duplicate enumerations. The person-place model includes a variable designed to have a higher value when a household member is found at more than one address across the AR sources used in creating the AR rosters (Morris 2017). The result is that the score assigned to the AR roster also has a higher value. Therefore, if a household includes a child or other member who appears at more than one address in the AR sources, the address is less likely to qualify to receive an AR enumeration. Such a household is more likely to be contacted by an interviewer during NRFU if it did not submit a self-response.

2.3.2 Corroboration requirement

An AR household roster for an address with a MAFID must meet the requirement that multiple sources indicate the family lives at the address. At least one person on the roster had to be corroborated at the address by two or more AR sources for the address to be eligible to be used in AR enumeration. A person-place (i.e., PIK-address) combination is considered corroborated if that person-place combination can be found in multiple AR sources. For this purpose, staff consider an agency or data supplier as a single source. For example, IRS 1040 and IRS 1099 files are treated as a single source since the two files come from the same agency. Thus, IRS 1099 cannot be used to corroborate a person from IRS 1040 (Mule, Keller, Konicki 2018).

Furthermore, the corroboration operation added AR sources to those used for constructing the AR household rosters. The additional sources include third-party sources, previous censuses and surveys such as the American Community Survey (ACS), and other federal sources such as the Selective Service System. The records used in the corroboration were maintained in a Census Bureau database known as PEARSIS (Production Environment for Administrative Records, Staging, Integration and Storage) that manages administrative records and services associated with administrative records (U.S. Census Bureau 2019). PEARSIS acquires data from other sources, both public and private, on an ongoing basis. The sources are listed below.

- USPS NCOA (U. S. Postal Service National Change of Address File)
- 2000 and/or 2010 CUF (Census Unedited File)
- Medicare
- HUD PIC (U.S. Department of Housing and Urban Development Public and Indian Housing Information Center)
- HUD TRACS (U.S. Department of Housing and Urban Development Tenant Rental Assistance)

- HUD CHUMS (U.S. Department of Housing and Urban Development Computerized Home Underwriting Management System)
- IHS (Indian Health Service) Patient file
- HUD PIC TRACS (U.S. Housing & Urban Development Public & Indian Tenant Rental Certification System)
- SSS (Selective Service System) Registration file
- Targus Wireless
- Targus Federal Consumer
- Targus Address File
- VSGI NAR (Veteran Service Group of Illinois Name and Address Resource file)
- VSGI Tracker/CRD (Veteran Service Group of Illinois Tracker Plus)
- DAR (Data. Analytics. Research) Partners

2.3.3 Identifying addresses for AR enumeration

2.3.3.1 Development of methodology

One of the goals during the planning of the 2020 Census was to reduce the cost of the NRFU fieldwork while maintaining the high quality of the household rosters used for census enumeration in NRFU. As part of the preparations for the 2020 Census, Census Bureau researchers developed a method for identifying addresses with high quality AR rosters suitable for enumeration. If a self-response was not submitted for an address but a high-quality AR roster was available, the address was removed from NRFU fieldwork, and the AR roster was inserted for the enumeration.

The approach for using ARs to enumerate some households in the 2020 Census uses a distance function approach based on two predictive models to identify addresses with a reliable household or a vacancy determination. These models are described in the next section.

The development of the methodology produced a screening tool that was flexible enough to permit adjustments of the threshold for identifying AR rosters suitable for enumeration if required by circumstances such as the cost or time constraints to complete NRFU fieldwork. One advantage of the approach was that the quality standard for the AR rosters could be raised or lowered if necessitated by circumstances arising from the status of the fieldwork or other situations that may occur in the census operations. The approach involved creating a housing unit level data set for the addresses and using characteristics of each address to assign a quality score for the AR roster for the address.

The initial work occurred during the 2015 census test and used the predictive modeling in an optimization approach to identify high quality AR rosters. The goal was to assign a rank to each AR roster that reflected its quality. The rankings would permit developing a cutoff that indicated AR rosters that were suitable for AR enumeration.

Subsequently, the researchers used 2010 Census data to compare the results from the distance function approach to the results from the optimization approach. The study found a high level of agreement between the quality assignments to the AR rosters by the two approaches. The study found that 91 percent of the addresses assigned the status of AR Vacant by the distance function approach also were assigned the AR Vacant by the linear optimization approach. A similar study

that compared the assignments of AR Occupied to the 2010 NRFU status found an agreement rate of 94.4 percent among the 500,000 addresses with the lowest scores, which indicated the highest probability, of being occupied.

In the end, the distance approach for assigning scores to the AR rosters for the addresses in NRFU was selected because it provided more flexibility and was easier to implement in the NRFU operation than the linear optimization approach.

2.3.3.2 Implementation of AR Modeling

Initially, a multinomial logistic regression model, known as the Vacant-Nonresidential model, is fit using data from 2010 Census to aid in identifying addresses (MAFIDs) that are vacant and nonresidential. The dependent variable has three levels, which are Vacant, Nonresidential, and Occupied. This model relies heavily on Undeliverable As Addressed (UAA) codes assigned by mail carriers for the U.S. Postal Service when they cannot deliver mail to an address, but also uses additional characteristics of the housing unit and its neighborhood from other data sources. The model is applied using the characteristics of the 2020 addresses (MAFIDs). The identification of AR Vacant and AR Nonresidential addresses relies on thresholds for the estimated probabilities of Vacant and Nonresidential identified using 2010 Census data. The addresses (MAFIDs) that the model does not identify as AR Vacant or AR Nonresidential are given the status of No Determination and retained for further processing.

Next, AR rosters are created for the addresses (MAFIDs) identified as occupied by the model discussed in the previous paragraph. Two models are fit using 2010 Census data and then applied to a distance measure that is used to identify the high-quality AR rosters for use in AR enumeration. One of the models is the Household Composition Model, which is used to estimate the probability that the household composition for the AR roster at an address (MAFID) is correct, say $\hat{p}_h^{HHC}$ where h is an index indicating the housing unit at the address.

The second model is a Person-Place model that provides an estimated probability that a person in the hth housing unit is correctly enumerated at the correct location $\hat{p}_{hi}^{person}$ where $i = 1,...n_h$ is an index for the people in the AR roster of size n_h. The person-place probability $\hat{p}_h^{person}$ assigned to housing unit h is the minimum value of the estimated probabilities $\hat{p}_{hi}^{person}$ over all the n_h people in the AR roster at the address (MAFID). Taking the minimum estimated person-place probability is a conservative approach.

A distance measure combines the predicted probabilities from the two models to create a single score that indicates the quality of the AR roster for an address (MAFID). The Person-Probability $\hat{p}_h^{person}$ that estimates the probability that the people on the AR roster are enumerated at the correct address and the Household Composition probability $\hat{p}_h^{HHC}$ combine into a single score, δ using a distance function. The score for address h is defined as follows:

$$\delta_h = \sqrt{\left(1 - \hat{p}_h^{person}\right)^2 + (1 - \hat{p}_h^{HHC})^2}.$$

The addresses assigned a score δ_h below a specified threshold are considered to have an AR roster of sufficient quality to be used in enumeration. Note that lower scores of δ_h indicate higher values of the estimated probabilities $\hat{p}_h^{person}$ and $\hat{p}_h^{HHC}$ and thereby, higher quality of the AR roster. Analyses with the 2010 Census data determined the threshold.

AR rosters that met the high-quality threshold required for AR enumeration also were required to have at least one person on the AR roster that appeared at the address in one of the corroboration sources listed in Section 2.3.2. AR rosters that met these criteria were called One-Visit Multiple Source AR rosters because they were eligible to be used to enumerate the address if a NRFU enumerator was unable to resolve the status after one visit and at least one person could be corroborated as residing at the address. Under these circumstances, the qualifying AR roster was inserted as the enumeration for the address.

2.3.4 NRFU Closeout AR statuses and corroboration

After the 2018 Census Test, which included a test of AR enumeration methodology, the staff identified a need for creating a set of AR Closeout statuses that relaxed the threshold used for the One-Visit Multiple Source AR rosters to help finish the NRFU operation. This strategy permitted AR rosters of slightly less quality than the One-Visit Multiple Source AR rosters to be used for AR enumeration. Fortunately, the distance function approach discussed in Section 3.3 and selected for the AR modeling had the flexibility to create the required AR Closeout statuses in the timeframe available to complete the census enumeration. The development of the AR Closeout statuses was based on research with 2010 Census data and used an approach similar to the methodology discussed in Section 3.3 for the assignment AR Occupied, AR Vacant, and AR Nonresidential. A study indicated the quality thresholds used for the One-Visit Multiple Source AR rosters could be increased somewhat and be used for AR enumeration when NRFU interviewers were unable to resolve the status of an address before the NRFU Closeout phase started (Keller 2019). The addresses assigned an AR Closeout status would not receive additional visits during NRFU Closeout that occurred at the end of NRFU when a specified percentage of the addresses in an area were resolved. However, when the NRFU operation entered its Closeout Phase, these addresses would be assigned their AR Closeout status and would not be reopened for additional contact attempts. The assignment of an AR Closeout roster to an address that received the AR Closeout Occupied status required corroboration of at least one person at the address by two sources.

In September 2020 as the AR processing entered its final stage, IRS approved the use of its data to determine the number of household members at a NRFU address when the household could not be corroborated with other data sources. IRS did not approve the Title 13 use of names or characteristics of the household members identified on IRS 1040s and 1099s when another corroboration source was not available. The use of only the household size for sole source corroboration was also applied to Medicare and Indian Health Service data. The rationale was that using only the household size did not reveal information that could lead to the identification of a person on the household roster. Using these sources to obtain a household size was preferred and provided higher quality data than could possibly be obtained through NRFU Closeout fieldwork or imputation. The rosters created by this procedure were called Household Size Only AR rosters.

In this analysis, when an address met the criteria for AR Occupied but the associated AR rosters formed using IRS data that could not be corroborated by another source, it received the designation "One-Visit Single Source AR Occupied." Similarly, the addresses that met the criteria for "AR Closeout Occupied" but had AR rosters that could not be corroborated by another source received the status "AR Closeout Occupied Single Source." During the NRFU operation, AR determinations based on a single source were only used in the closeout phase.

2.4 Research questions

The goal of our research is to gain a better understanding of why some self-responses and NRFU household member responses have household sizes that differ from the roster created using ARs for the corresponding address. The focus is on the situations where the self-response roster size differs from the AR roster size. For the situation where the self-response or NRFU household member response has one less person, that additional person must appear on an AR delivered to the Census Bureau between May but no later than August 2020. The address associated with the AR roster also must have a Master Address File identification number (MAFID) assigned by the Census Bureau. In this document, addresses with AR rosters are assumed to have a MAFID.

The research seeks to answer the following questions:

1) For addresses where the census household size and the AR household size are different, what is the distribution of the differences in household size? Do the distributions of the differences in household size vary by type of AR roster and/or by type of respondent? If so, how are the distributions different?
2) What are the characteristics of the households that tend to have a census household size that is larger than the AR household size for the address?
3) When the census roster and the AR roster for an address differ by one person, what are the characteristics of the person found on one roster but not the other?

3 Methodology

3.1 Analysis approach

The analyses begin by examining cross-tabulations of counts of self-response (SR) rosters by administrative records (AR) roster for addresses where the SR roster count disagrees with the AR roster. Next, the analyses examine cross-tabulations of the household compositions of self-response (SR) roster counts by the household compositions of administrative records (AR) roster counts.

The analyses consider two types of responses to the census, which are self-responses and NRFU interviews where the respondent was a household member. These types of responses are believed to be of the highest quality because the respondent resides at the address.

- Self-response: A household could submit a census response by internet, mail, or telephone. The self-response period began prior to the decision to delay the data

collection but was extended beyond the planned end date. Self-response was scheduled from March 12 to July 31 but was conducted from March 12 to October 15.

- NRFU Household Member response: When the Census Bureau did not receive a self-response, NRFU enumerators visited address to obtain an interview that would resolve the status of the address and enumerate the residents when occupied. The start of the NRFU operation was delayed and the operation was extended. NRFU was originally scheduled from May 13 to July 24 but was conducted from July 16 (soft launch) through October 15.

3.2 AR rosters Used in Final Census Enumeration

Before beginning the discussion of the comparisons of AR rosters and census rosters for addresses that had both, Table 3.1 provides some perspective on the scope of the use of AR rosters for census enumeration. The table also sheds some light on the quality of the AR rosters that were used to enumerate some of the addresses in NRFU since neither was a self-response submitted nor did fieldwork resolve the status of the address. Table 3.1 shows that 91.95 percent of the AR rosters used for enumeration were in the category One-Visit Multiple Sources, which contains the AR rosters of the highest quality.

Table 3.1: Administrative Record Occupied Housing Units by Roster Quality Category

AR Roster Quality Category	Count	Percent
One-Visit multiple sources	4,468,000	91.95%
Closeout multiple source	186,000	3.83%
Closeout Household size only	188,000	3.87%
Other	16,500	0.34%
Total	4,859,000	100.00%

Note: Closeout Household Size Only combines One-Visit Single Source and Closeout Occupied Single Source. Counts are rounded for disclosure avoidance.

4. Results

Section 4 focuses on comparisons of the household size and characteristics of census rosters and AR rosters at addresses where both types of rosters are available, and each has at least one person.

In particular, this section concentrates on addresses where the AR and census rosters differ on the number of household members. The topics include the following:

- Section 4.1 contains results aggregated at the national level that focus on the difference in household size observed for the AR roster and the census roster.
- Section 4.2 shows the frequency of roster characteristics when the two types of rosters differ on the size of the household.

- Section 4.3 focuses on addresses where the size of the census roster and the AR roster differ by one in its analyses. This includes determining whether the people found on one roster but not the other have an AR in the file created for census quality assurance called the Self-response Verification file (Deaver 2020). In addition, if the people were found at the address, the study examines when they were found, meaning whether ARs occurred before Census Day, after Census Day, or both before and after.
- Section 4.4 contains additional analysis of age and relationship of people found on only one of the rosters. This subsection looks at the distribution of the age of the person found on one but not both rosters. Since relationship is reported and imputed when missing for the Census, this subsection also looks at the relationship of the extra census person to the householder.

4.1 Aggregate and household size comparison results

Table 4.1.1 shows national level results for the total number of addresses and the total number of those addresses that had an AR roster of high enough quality to be suitable for census enumeration. Also shown are the number of Occupied addresses with either a self-response or a NRFU household member response that had a high-quality AR roster suitable for enumeration.

Table 4.1.1. Total Addresses vs Addresses with an Administrative Record Roster

	U.S. Total (count)	Percent of Total Addresses
U.S. Overall		
Total addresses	151,800,000	100.0%
Total addresses with an administrative record roster	65,850,000	43.4%
Final Status of Occupied Addresses:		
Self-Response	97,580,000	64.3%
Self-Response with an administrative record roster	55,350,000	36.5%
NRFU Household Member Response	16,460,000	10.8%
NRFU Household Member Response with an administrative record roster	3,781,000	2.5%

The rest of the analyses in this section will focus on Occupied addresses with either a self-response or a NRFU household member response that also had a high-quality AR roster that qualified for census enumeration. In Table 4.1.2 the column "same count" displays the percentage of addresses where the administrative record (AR) household size is equal to the census response household size. The column "different count" displays the percentage of addresses where the administrative record household size is different from the census response household size. The range of the household size agreement rates for the three self-response

modes, internet, mail, and telephone, is 79.4 percent to 80.1 percent. Section 5 contains a discussion of plausible explanations of reasons why a person may be found on one of the rosters for an address but not the other.

Table 4.1.2. Agreement rate between AR household size and census self-response household size by mode of self-response

Self-response Mode	Same count (%)	Different count (%)	Total (%)	Total count
Internet	79.4	20.6	100.0	43,910,000
Mail	80.1	19.9	100.0	10,440,000
Telephone	80.1	19.9	100.0	1,003,000
Modes Combined	79.5	20.5	100.0	55,350,000

Note: Mail includes Mailout/Mailback and Update/Leave

In Table 4.1.3 the modes of self-response are aggregated for a comparison with NRFU household member responses. Note that 79.5 percent of the census rosters from self-responses have a household size that agrees with the AR roster for the same address. The agreement rate for household size between NRFU household member responses and their corresponding AR roster is 58.7 percent, which is 20.8 percentage points lower than the 79.5 percent agreement rate observed between self-response rosters and their corresponding AR rosters.

Table 4.1.3. Agreement rate between AR household size and Self-response or NRFU household size

Address status	Same count (%)	Different count (%)	Total (%)	Total count
Self-response	79.5%	20.5%	100%	55,350,000
NRFU Household Member Response	58.7%	41.3%	100%	3,781,000

Table 4.1.4 expands on Table 4.1.3 by showing the size of the difference in the household size between the census roster and the AR roster. A negative difference means that the AR roster has more people than the census household roster. However, if the census roster is larger, then the difference is positive. The distribution for the difference in the census roster and AR roster size for the self-responses is reasonably symmetrical. The distribution of the difference in the census roster and AR roster for the NRFU household interviews is slightly less symmetrical.

Table 4.1.4. Percentage of Self-responses and NRFU household member responses by the difference between the 2020 Census household size and the AR household size

Address status	Household size difference (Census HH size – AR HHsize)							Total	Total count
	<=-3	-2	-1	0	1	2	3>=		
Self-response	0.4%	0.9%	8.1%	79.5%	8.6%	1.7%	0.8%	100%	55,350,000
NRFU Household Member Response	1.9%	3.0%	12.7%	58.4%	15.8%	4.8%	3.3%	100%	3,781,000

Note: Self-response includes internet, mail, and telephone responses.

Table 4.1.5 provides more detail about the distribution of the differences between the census roster size and the AR roster size by type of response and two other criteria. One criterion that indicates the quality of the AR roster is whether the value of the distance function met the threshold for only one visit in NRFU or the higher threshold for inclusion in the NRFU Closeout phase. Another criterion is whether the census address was resolved by July 30, 2020, which indicates how close to Census Day the data used in creating the AR roster was collected.

The large majority of the NRFU household member responses were collected after July 30. For all three categories of AR rosters, the self-responses resolved by July 30 have 83.8 percent agreement on household size between the census roster and the AR roster. The self-responses unresolved on July 30 have an agreement rate of 74.4 percent, and the NRFU Household Member responses have an agreement rate of 66.3 percent. For the Closeout Multiple Source AR rosters, the percentages of addresses where the AR roster has one more person than the census roster range from 24.9 percent to 29.1 percent across the three time periods. The Closeout Multiple Source AR rosters were permitted to qualify for use with a larger distance measure than the One-Visit Multiple Source AR rosters, as discussed in Section 4.3.5. The larger distances often were because of the AR source having the person at multiple addresses. If the census roster for the address did not include that AR person, a possible explanation is that the AR roster had a higher count than census for these addresses.

Table 4.1.5. Difference in census household size and AR household size by AR roster type, census response type, and whether address status was resolved by July 30.

AR roster distance function and corroboration category by Response Type and whether resolved by July 30	Household size difference (=Census HH size – AR HH size)								Total count
	AR Higher			Same	Census Higher				
	<=-3	-2	-1	0	1	2	3>=	Total	
One-visit multiple source (2.2 people per HH)									
Self-Response									
Resolved by July 30	0.2%	0.5%	6.1%	83.8%	7.4%	1.3%	0.6%	100.0%	44,540,000
Unresolved by July 30	0.6%	1.0%	8.3%	74.4%	11.7%	2.6%	1.4%	100.0%	1,957,000
NRFU HH member	1.4%	2.1%	10.7%	66.3%	13.7%	3.6%	2.2%	100.0%	2,083,000
Closeout multiple source (2.9 people per HH)									
Self-Response									
Resolved by July 30	1.8%	3.9%	29.1%	51.6%	10.0%	2.4%	1.3%	100.0%	4,431,000
Unresolved by July 30	2.7%	5.2%	25.4%	46.9%	13.6%	3.9%	2.3%	100.0%	305,000
NRFU HH member	3.9%	6.7%	24.9%	41.3%	14.6%	5.0%	3.5%	100.0%	647,000
Closeout Household Size Only (1.9 people per HH)									
Self-Response									
Resolved by July 30	0.4%	0.9%	5.3%	68.3%	17.7%	4.6%	2.9%	100.0%	3,682,000
Unresolved by July30	0.7%	1.3%	6.7%	64.6%	18.3%	5.1%	3.5%	100.0%	429,000
NRFU HH member	1.6%	2.7%	9.0%	53.3%	20.8%	7.1%	5.5%	100.0%	1,052,000

Note: Resolved by July 30 means a response was submitted by July 30 and Unresolved by July 30 means that a response was submitted after July 30. Census nationwide launch date of Nonresponse Followup (NRFU) operations was August 11.

4.2 Characteristics of census households when census and AR rosters differ

This section examines the characteristics of the census rosters at addresses where the household size for the census roster differs from the household size for the AR roster. The analyses in this section build on Table 4.1.5 by examining the traits of the census rosters at addresses where the census roster has more people than the AR roster for each of three types of AR rosters.

Table 4.2.1 shows the percentages for several amounts of household size disagreement between AR rosters and census self-responses and NRFU household member responses for the same addresses. The analyses in this section are descriptive and not for making inferences but instead aimed at providing guidance on where to focus further research. The analyses include an examination of characteristics of the household members when the size of the census roster and the size of the AR roster disagree by one person for three types of AR rosters: The three types are One-Visit Multiple Source AR Rosters in Table 4.2.1, Closeout Multiple Source AR Rosters in Table 4.2.2, and Closeout Household Size Only AR Rosters in Table 4.2.3. Also included in

the tables is the percentage of the addresses where the census roster is larger than the AR roster. Appendix A contains the results found in Tables 4.2.1, 4.2.2, and 4.2.3 in one table to facilitate comparisons between the three types of rosters.

Of particular interest in this section are addresses where the census self-response roster size and the AR roster size differ by one. Note that Table 4.1.5 in the previous section shows that when the census roster and AR roster for an address did not agree on size, the large majority differed by one. This phenomenon was not observed during the research and preparations for AR enumerations. The final refinements to the methodology for forming AR rosters were based on analysis of the 2018 End-to-End Census Test using the June delivery of data from IRS 1040s and 1099s.

Even if a relatively large number of addresses in the 2018 test appeared to have census rosters and AR rosters that differed by one person and there had been research on this topic, the effect of the pandemic could not have been anticipated and incorporated in the analyses. In the summer of 2020, staff recognized there were a substantial number of addresses where the AR and census roster sizes differed by one, but there was not enough time to do the research needed to formulate rules to implement a remedy and meet the deadline for the initial releases of 2020 Census data.

4.2.1 One-Visit Multiple Source AR rosters

Table 4.2.1 contains categories that describe traits of one or more members of the household listed on a census roster for addresses with a One-Visit Multiple Source AR roster where the sizes of the census and AR rosters differ by one. The census roster may be from a self-response or NRFU household member response. The responses were combined to provide more data for this analysis. Each category contains levels of the trait along with the number and percentage of the total households for each level. Also, the percentages for the categories within a trait do not necessarily add to 100 percent.

Table 4.2.1: Census roster traits for addresses with a census self-response or a NRFU household member response roster in addition to a <u>One-Visit Multiple Source</u> AR roster where the sizes of the two rosters differ by one

Census Roster Trait	Total	Column Percent	Percent Census is higher when roster sizes are different
Overall	8,397,000	100.0%	
NH White Alone Householder	5,680,000	67.6%	55.2%
NH Black Alone Householder	911,000	10.9%	60.9%
NH AIAN Alone Householder	32,500	0.4%	61.9%
NH Asian Alone Householder	432,000	5.1%	62.1%
NH NHPI Alone Householder	7,400	0.1%	64.1%
NH SOR Alone Householder	35,500	0.4%	59.6%
NH 2 or More Races Householder	299,000	3.6%	60.7%
Hispanic Householder	1,000,000	11.9%	66.4%
At least One NH White Alone Person	6,027,000	71.8%	56.7%
At least One NH Black Alone Person	1,036,000	12.3%	64.0%
At least One NH AIAN Alone Person	57,500	0.7%	71.9%
At least One NH Asian Alone Person	551,000	6.6%	66.4%
At least One NH NHPI Alone Person	14,500	0.2%	73.6%
At least One NH SOR Alone Person	62,500	0.7%	70.6%
At least One NH 2 or More Races Person	599,000	7.1%	69.3%
At least One Hispanic	1,347,000	16.0%	69.9%
Spouse Present	3,605,000	42.9%	73.3%
Unmarried Partner Present	785,000	9.3%	87.6%
Children Present	3,642,000	43.4%	75.1%
Other Relatives Present including Grandchildren	1,247,000	14.9%	93.5%
Non-Relatives Present	710,000	8.5%	94.8%
At least one person age 0 to 4	967,000	11.5%	82.2%
At least one person age 5 to 17	2,008,000	23.9%	64.4%
At least one person age 18 to 24	1,346,000	16.0%	85.2%
At least one person age 25 to 49	4,511,000	53.7%	69.2%
At least one person age 50 to 64	3,769,000	44.9%	62.0%
At least one person age 65+	2,616,000	31.2%	61.6%

Note: Source of characteristics is the census roster.

The last column contains the percentage of the addresses where the census roster has one more person than the AR roster within a census roster trait. Notice that for each category, the percentage of addresses with the census roster larger than the AR roster is greater than 50 percent for all the listed traits. These percentages range from 55.2 percent to 94.8 percent. However, the

range of the number of addresses within the different roster trait categories is 14,500 to 3,769,000.

Two types of traits of the household members have categories that cover all the households where the census and AR rosters have household sizes that differ by one. These traits also have categories that are large enough for research aimed at identifying the underlying reasons for the difference in the size of the census roster and the AR roster.

One of these traits is the relationship of a household member to the householder. This trait has six relationship categories with sizes that range from 710,000 to 3,642,000. The percentage of the addresses in the relationship categories where the census roster is larger than the AR roster ranges from 73.3 percent to 94.8 percent.

The other trait of interest is age, which is divided into six age ranges with sizes ranging from 967,000 to 4,511,000. The percentage of the addresses in the age categories where the census roster is larger than the AR roster ranges from 61.6 percent to 85.2 percent.

4.2.2 Closeout Multiple Source AR rosters

In this section, the investigation turns to the Closeout Multiple Source AR rosters at addresses that also have a census self-response roster or a NRFU household member response roster where the sizes of the two rosters differ by one person. Table 4.2.2 shows results in the same manner as Table 4.2.1 for the One-Visit Multiple Source AR rosters but the data is for the Closeout Multiple Source AR rosters. The traits and categories within the traits in Table 4.2.2 are the same as in Table 4.2.1. As with the One-Visit Multiple Source AR rosters discussed in Section 4.2.1, the concern is whether the categories within the traits shown in Table 4.2.2 are large enough for analyses that will examine underlying reasons for the difference in the census and AR roster sizes for the addresses. Also, of interest for the analyses is quantifying the difference incurred by Closeout Multiple Source AR rosters needing to meet a slightly lowered quality criteria than required of the One-Visit Multiple Source AR rosters.

Table 4.2.2 shows that the number of addresses with Closeout Multiple Source AR rosters is 2,687,000, which is about 32 percent of the number of addresses with One-Visit Multiple Source AR rosters shown in Table 4.2.1. The percent of the addresses in the trait groups where the census roster is larger than the AR roster ranges from 24.6 percent to 74.2 percent. Both the size of the trait groups and the percentage of addresses in the groups that have a census roster larger than the AR roster tend to be substantially smaller than observed for the One-Visit Multiple Source AR rosters.

Table 4.2.2. Census roster traits for addresses with a census roster from a self-response or a NRFU household member response plus a <u>Closeout Multiple Source</u> AR roster where the sizes of the two rosters differ by one

Census Roster Trait	Total	Column Percent	Percent Census is higher when roster sizes are different
Overall	2,687,000	100.0%	
NH White Alone Householder	1,712,000	63.7%	24.6%
NH Black Alone Householder	348,000	13.0%	34.4%
NH AIAN Alone Householder	13,000	0.5%	41.2%
NH Asian Alone Householder	131,000	4.9%	43.1%
NH NHPI Alone Householder	2,800	0.1%	47.0%
NH SOR Alone Householder	11,500	0.4%	38.2%
NH 2 or More Races Householder	94,000	3.5%	33.5%
Hispanic Householder	374,000	13.9%	46.8%
At least One NH White Alone Person	1,810,000	67.4%	25.9%
At least One NH Black Alone Person	382,000	14.2%	36.5%
At least One NH AIAN Alone Person	20,500	0.8%	46.7%
At least One NH Asian Alone Person	160,000	6.0%	44.3%
At least One NH NHPI Alone Person	4,700	0.2%	51.7%
At least One NH SOR Alone Person	19,500	0.7%	47.3%
At least One NH 2 or More Races Person	183,000	6.8%	41.8%
At least One Hispanic	474,000	17.6%	47.5%
Spouse Present	1,236,000	46.0%	39.7%
Unmarried Partner Present	191,000	7.1%	50.6%
Children Present	1,140,000	42.4%	54.1%
Other Relatives Present including Grandchildren	314,000	11.7%	74.2%
Non-Relatives Present	144,000	5.4%	73.9%
At least one person age 0 to 4	306,000	11.4%	61.5%
At least one person age 5 to 17	799,000	29.7%	52.9%
At least one person age 18 to 24	390,000	14.5%	64.7%
At least one person age 25 to 49	1,418,000	52.8%	44.2%
At least one person age 50 to 64	1,108,000	41.2%	31.2%
At least one person age 65+	736,000	27.4%	25.1%

Note: Source of characteristics is the census roster.

4.2.3 Closeout Household Size Only AR Rosters

In this section, the attention focuses on the Closeout Household Size Only AR Rosters at addresses that also have a census roster and the number of people listed on the two rosters differ by one. The number of addresses in this group is 1,813,000, which is about 21.6 percent of the number of One-Visit Multiple Source AR rosters shown in Table 4.2.1 and 67 percent of the number of Closeout Multiple Source AR rosters shown in Table 4.2.2. For the Closeout Household Size Only AR Rosters, the observed percentages of addresses with a census roster larger than the AR roster when the sizes differ by one range from 71.0 percent to 96.0 percent across the categories within the roster traits. Both the size of the trait groups and the percentage of addresses in the groups that have a census roster larger than the AR roster tend to be considerably lower than observed for the One-Visit Multiple Source AR rosters.

Table 4.2.3: Census roster traits for addresses with a census roster from a self-response or a NRFU household member response plus a Closeout Household Size Only AR roster where the sizes of the two rosters differ by one

Census Roster Trait	Total	Column Percent	Percent Census is higher when roster sizes are different
Overall	1,813,000	100.0%	
NH White Alone Householder	836,000	46.1%	77.0%
NH Black Alone Householder	302,000	16.7%	71.0%
NH AIAN Alone Householder	10,500	0.6%	75.4%
NH Asian Alone Householder	139,000	7.7%	81.1%
NH NHPI Alone Householder	3,700	0.2%	74.2%
NH SOR Alone Householder	10,500	0.6%	77.4%
NH 2 or More Races Householder	77,500	4.3%	77.1%
Hispanic Householder	433,000	23.9%	79.1%
At least One NH White Alone Person	936,000	51.6%	78.7%
At least One NH Black Alone Person	344,000	19.0%	73.7%
At least One NH AIAN Alone Person	17,000	0.9%	81.2%
At least One NH Asian Alone Person	166,000	9.2%	83.8%
At least One NH NHPI Alone Person	5,600	0.3%	78.9%
At least One NH SOR Alone Person	19,000	1.0%	82.8%
At least One NH 2 or More Races Person	144,000	7.9%	82.0%
At least One Hispanic	523,000	28.9%	80.9%
Spouse Present	580,000	32.0%	89.5%
Unmarried Partner Present	341,000	18.8%	92.8%
Children Present	803,000	44.3%	83.9%
Other Relatives Present including Grandchildren	302,000	16.7%	94.3%
Non-Relatives Present	311,000	17.2%	96.0%
At least one person age 0 to 4	333,000	18.4%	84.4%
At least one person age 5 to 17	516,000	28.5%	82.2%
At least one person age 18 to 24	534,000	29.5%	88.5%
At least one person age 25 to 49	1,314,000	72.5%	79.0%
At least one person age 50 to 64	499,000	27.5%	80.2%
At least one person age 65+	200,000	11.0%	83.6%

Note: Source of characteristics is the census roster.

4.3 Census and AR rosters differ by one

The investigation in this section builds on the results of Section 4.2 by using only the One-Visit Multiple Source AR rosters shown in Table 4.2.1 since these AR rosters have an adequate number of observations to support further analyses while the other two types of AR rosters do not. An additional criterion for an address to be included in the analyses in this section is that the people on both rosters can be assigned a PIK, which enables identifying the additional person on either side. The requirement also enables examining whether the person can be found at the address of the enumeration in the Self-response Verification file (Deaver 2020), which is a composite file created specifically for the census quality assurance operation. The advantage of the Self-response Verification file is that it enables determining whether the person had ARs at one or more other addresses in the Self-response Verification file before Census Day, after Census Day, or both before and after Census Day.

In this section, the focus is on addresses where the census roster and the AR roster differ by one person for census self-responses and NRFU household member responses. The analyses do not include addresses where the census roster and the AR roster differ by more than one person because there are not enough observations for this type of analysis as shown in Table 4.1.5. Section 5 contains a discussion of plausible explanations of reasons why a person may be found on one of the rosters for an address but not the other.

The analyses divide the addresses where the AR roster and the census roster differ by one person into two groups, one where the AR roster has one more person than the census roster and the other where the AR roster has one less person. The study examines whether the person on one of the rosters, but not on both can be found at the address in the Self-response Verification file at the same address or a different address and if found, whether the timing is before, after or including (i.e., before and after) Census Day. In addition, the analyses examine the characteristics of the people on one of the rosters but not both.

4.3.1 AR roster has one <u>more</u> person than census roster

This section focuses on addresses that were either census self-responses or NRFU household member responses where the One-Visit Multiple Source AR roster had one <u>more</u> person than the census roster. The analysis considers whether the extra person was found at the same address or a different address in the Self-response Verification file and the timing of the records for the person. Also considered is whether the extra people had records in the Self-response Verification file before, after, or including (i.e., before and after) Census Day.

Table 4.3.1 contains the distribution of where a person included on the One-Visit Multiple Source AR roster for an address but not the census roster was found in the Self-response Verification file by whether the location was the same address or a different address and the timing of the records relative to Census Day. Among census self-responses with a person on the One-Visit Multiple Source AR roster but not the census roster, 97.98 percent were found at the same address in the Self-response Verification file while 2.02 percent were found only at a different address. The percentages are comparable for NRFU household member responses with 97.48 percent found at the same address and 2.35 percent found at a different address.

Table 4.3.1. For One-Visit Multiple Source AR rosters with one <u>more</u> person than the self-response or NRFU HH member response for the address, the percentage of the extra people found at the same address in the Self-response Verification file, and the percentage found at a different address with the time period when their records were found.

Status in AR Composite File	AR record timing	Type of census response	
		Self-response % of total	NRFU HH member % of total
Found at same address	Before Census Day	15.92	17.65
	Includes Census Day	79.40	76.05
	After Census Day	2.64	3.36
	subtotal	97.98	97.48
Found at different address	Before Census Day	1.16	1.43
	Includes Census Day	0.75	0.84
	After Census Day	0.10	0.13
	subtotal	2.02	2.35
Combined	Total	100.00	100.00

Note: Percentages may not add to 100 because of rounding.

Table 4.3.2 focuses on people included on the One-Visit Multiple Source AR roster but not the census roster for an address and the time period when the records for individuals were found in the Self-response Verification file at a different address relative to Census Day. In addition, the table shows that the percentage of the extra people on self-response rosters found at another address before Census Day was 48.87 percent while the percentage of those found at another address during a time period that includes Census Day was 32.17 percent. Combining the Before and Includes Census Day categories accounts for 81.04 percent of the addresses with a self-response roster and an extra person on the One-Visit Multiple Source AR roster.

Table 4.3.2 also shows results for addresses with a NRFU household member response and a One-Visit Multiple Source AR roster where the AR roster has one more person. Among these addresses, 54.00 percent of extra people have a record in the Self-response Verification file at a different address in the time period before Census Day. In addition, 28.89 percent of the extra people on an AR roster have a record in the Self-response Verification file at a different address in a time period that includes Census Day. Combining these two categories indicates that 83.39 percent of the extra people at these addresses were found at another address either before Census Day or in a time period that includes Census Day. The implication of these results is that a person on the Multiple Source AR roster but not the census roster for the address with a census self-response or NRFU household member response is likely to have an AR(s) at another address close to Census Day.

Table 4.3.2. For One-Visit Multiple Source AR rosters with one <u>more</u> person than the self-response or NRFU HH member response- for the address and the extra person appearing in the Self-response Verification file at the same address in a time period <u>including Census Day</u>, the percentage of the extra people also found at another address by time period of the record relative to Census Day.

Timing of record(s) at another address	Self-response		NRFU HH response	
	% of total	% of category	% of total	% of category
Before Census Day	38.82%	48.87%	41.95%	55.00%
Includes Census Day	25.55%	32.17%	22.03%	28.89%
After Census Day	4.01%	5.05%	3.81%	5.00%
Not found at another address	11.05%	13.92%	8.90%	11.67%
Column Total	79.43%	100.00%	76.69%	100.00%

4.3.2 AR roster has one <u>less</u> person than census roster

Next, we consider the situation where the One-Visit Multiple Source AR roster has one <u>less</u> person than the census roster submitted by a self-response or a NRFU HH member response. Table 4.3.3 contains the distribution of both types of responses by whether the extra person was found at the same address or a different address in the Self-response Verification file and the timing of the record relative to Census Day, meaning before, after, or both before and after Census Day.

Table 4.3.3 shows that 74.93 percent of the people on a self-response roster but not the AR roster were found in the Self-response Verification file at the same address and 25.07 percent were found at a different address. The 74.93 percent found at the address is the sum of the 36.91 percent with only records before Census Day, the 34.63 percent with records including Census Day, and 3.38 percent with only records after Census Day. The 25.07 percent found at a different address is the sum of the 7.87 percent with only records before Census Day, the 16.33 percent with records including Census Day, and 0.87 percent with only records after Census Day. For addresses where the One-Visit Multiple Source AR roster had one less person than the census roster submitted in a NRFU response by a household member, 72.46 percent were found at the same address and 27.54 percent were found at a different address. These percentages for the NRFU household member responses are very similar to the percentages observed for extra people on the AR rosters for the self-responses.

Table 4.3.3. For One-Visit Multiple Source AR rosters with one <u>less</u> person than the self-response or NRFU HH member response for the address, the percentage of the extra people found at the same address in the Self-response Verification file, the percentage found at a different address and the timing of the records

Status in AR Composite File	Time found	Type of response	
		Self-response % of total	NRFU HH member % of total
Found at same address	Before Census Day	36.91	41.53
	Includes Census Day	34.63	27.54
	After Census Day	3.38	3.47
	subtotal	74.93	72.46
Found at different address	Before Census Day	7.87	10.17
	Includes Census Day	16.33	16.53
	After Census Day	0.87	0.93
	subtotal	25.07	27.54
Combined	Total	100.00	100.00

Note: Percentages may not add to 100 because of rounding.

Next, the focus turns to self-responses and NRFU household member response that had one less person than the One-Visit Multiple Source AR roster for the address, and the Self-response Verification file shows the extra person at the address on dates that include Census Day. For the self-response rosters, Table 4.3.4. shows that 48.18 percent of these people have AR records at another address on dates before Census Day and 32.07 percent were found at another address on dates including Census Day. Combining the categories for before and including Census Day accounts for 81.25 percent of the people found at another address. An additional 2.96 percent appear in the Self-response Verification file at another address after Census Day. However, 15.68 percent of the extra people were not found at another address in the AR composite.

Table 4.3.4 also shows comparable information for the NRFU responses by household members that reported one more person than the One-Visit Multiple Source AR roster. In addition, the Self-response Verification file indicates 32.31 percent of the extra people at the address on dates that include Census Day. The Self-response Verification file also shows 52.31 percent of these people at another address on dates before Census Day and an additional 2.77 percent appear in the Self-response Verification file at another address after Census Day. In addition, 13.54 percent of the extra people were not found at another address in the Self-response Verification file. As shown Tables 4.3.1 and 4.3.2 for the addresses where the AR roster has one more person than the census roster, a person on the AR roster but not the census roster for an address is likely to have an AR(s) at another address close to Census Day. In Table 4.3.4, the distributions of the timing of the record(s) for the self-responses and the NRFU HH member responses have a similar pattern. These two types of responses are of highest quality so possibly the similarity in the distributions is to be expected.

Table 4.3.4. For self-responses and NRFU HH responses with one <u>less</u> person than the One-Visit Multiple Source AR roster and the extra person appearing in the Self-response Verification file at the same address at a time <u>including Census Day</u>, the percentage of the extra people also found at another address by time period.

| Timing of record(s) at another address	Self-response		NRFU HH response	
	% of total	% of category	% of total	% of category
Before Census Day	16.69%	48.18%	14.29%	52.31%
Includes Census Day	11.46%	32.07%	8.82%	32.31%
After Census Day	1.02%	2.96%	0.76%	2.77%
Not found at another address	5.43%	15.68%	3.70%	13.54%
Column total	34.65%	100.00%	27.31%	100.00%

Note: Total for column "% of total" equals percentage for "Includes Census Day" row in Table 4.3.3.

4.4 Age of person when census roster and AR roster differ by one

The first step is to examine the age of the person who causes the census self-response roster and the One-Visit Multiple Source AR roster to differ by one. As with the previous analyses, the approach to the investigation divides the data by whether the AR roster has one more person or one less than the census roster. Plots for addresses with NRFU household member responses and AR rosters are not included in the line plots of the age distributions because the number of observations is insufficient for this type of analysis. However, these addresses are included in bar charts of the age distributions that include both self-responses and NRFU household member responses. Section 5 contains a discussion of the results.

### 4.4.1 **Age distribution when AR roster has one <u>more</u> person than census roster**

Figure 4.4.1 shows the age distribution for people on a One-Visit Multiple Source AR roster but not on the census roster for addresses that submitted a self-response. These age distributions reflect whether one or more records for the person were found at the address in the Self-response Verification file before, after, or including Census Day. In Figure 4.4.1, ages 86 and over have been truncated to age 85 for disclosure avoidance.

Figure 4.4.1: Age Distribution for Additional AR Person Not Found on Self-Response Census Roster
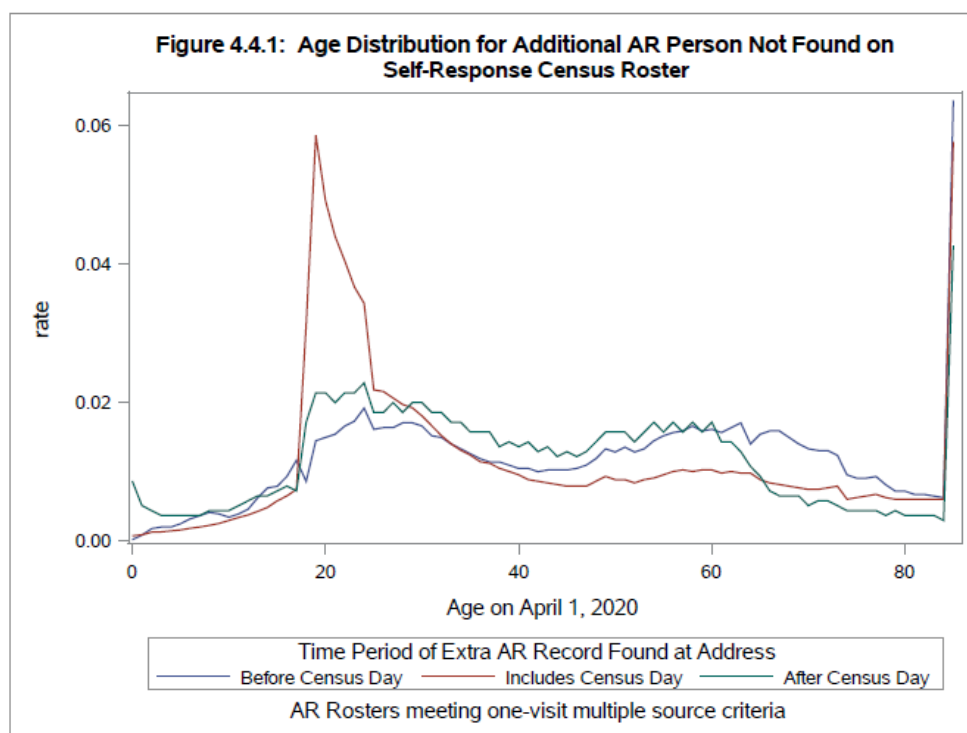
Note: Ages over 85 truncated at 85 for disclosure avoidance.

Apparently, the biggest spike in the distribution occurs at ages 18 to 29 when the person included on the AR roster but not the census roster was found at the address in the Self-response Verification file on one or more records that included Census Day. The group aged 18 to 29 is about 39 percent of the people found on the AR roster but not the census roster for an address with the bulk of the group in the 18-to-24 age range.

Figure 4.4.2 uses bar charts to display the distribution of age groups for the people on the AR rosters for an address but not on the census roster for NRFU household member responses in addition to the age distribution for the same type of people on self-response rosters. The number of NRFU household member responses is small enough to require bar charts where the ages can be aggregated to meet the requirements for disclosure avoidance. Therefore, aggregating the ages for self-responses is required for viewing the results of the two types of responses side-by-side. The age distributions for self-responses and NRFU household member responses are shown by whether records for the additional AR people were found in the Self-response Verification file before, after, or including Census Day. Displaying these plots for both types of responses side-by-side facilitates a visual comparison of the patterns in the age distributions of the people on the AR rosters but not the census rosters.

.

Figure 4.4.2: Age Distribution for Additional AR Person Not Found on Census Enumeration Roster
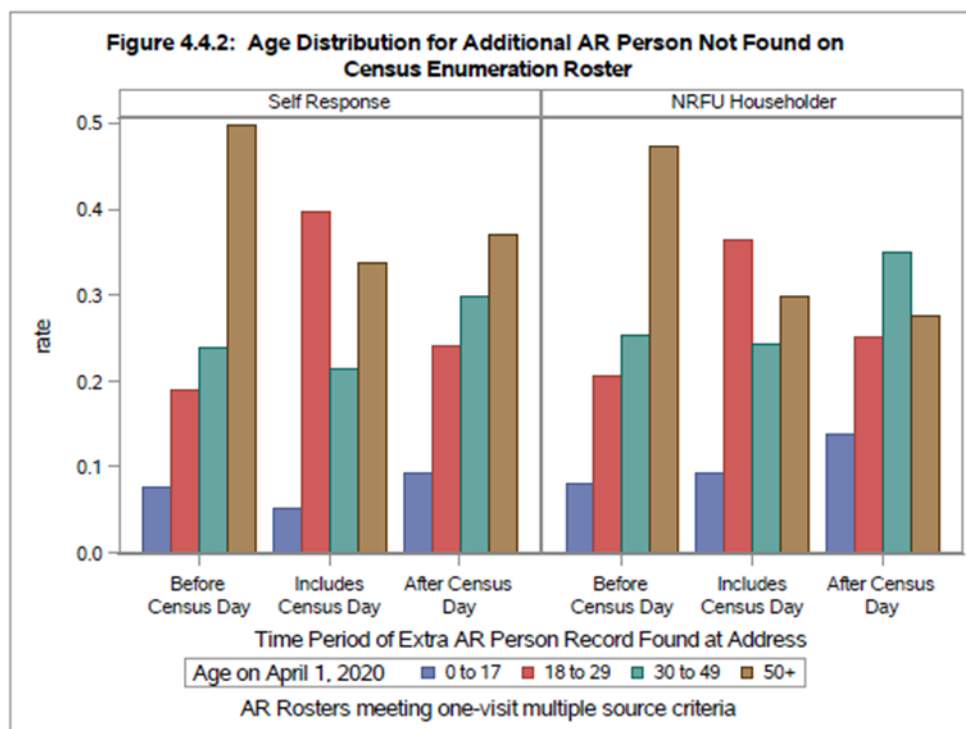
The age distributions for self-responses and NRFU household member responses in Figure 4.4.2 have similar patterns across the time periods. When there was a self-response for an address, the age distributions for the people found in Self-response Verification file at the address on a date or dates before Census Day and including Census Day appear similar to the corresponding age distributions for a NRFU HH member response. However, people found at the address in the Self-response Verification file only after Census Day have age distributions for self-responses and NRFU household member responses that appear less similar, but statistical tests have not been conducted.

4.4.2 Age distribution when AR roster has one <u>less</u> person than census roster

Figure 4.4.3 shows the age distribution for the person on the census roster but not on AR roster for addresses that submitted a self-response by whether a record for the person was found in the Self-response Verification file before, after, or including Census Day.

Figure 4.4.3: Age Distribution for Additional Self-Response Census Person Not Included on AR Roster
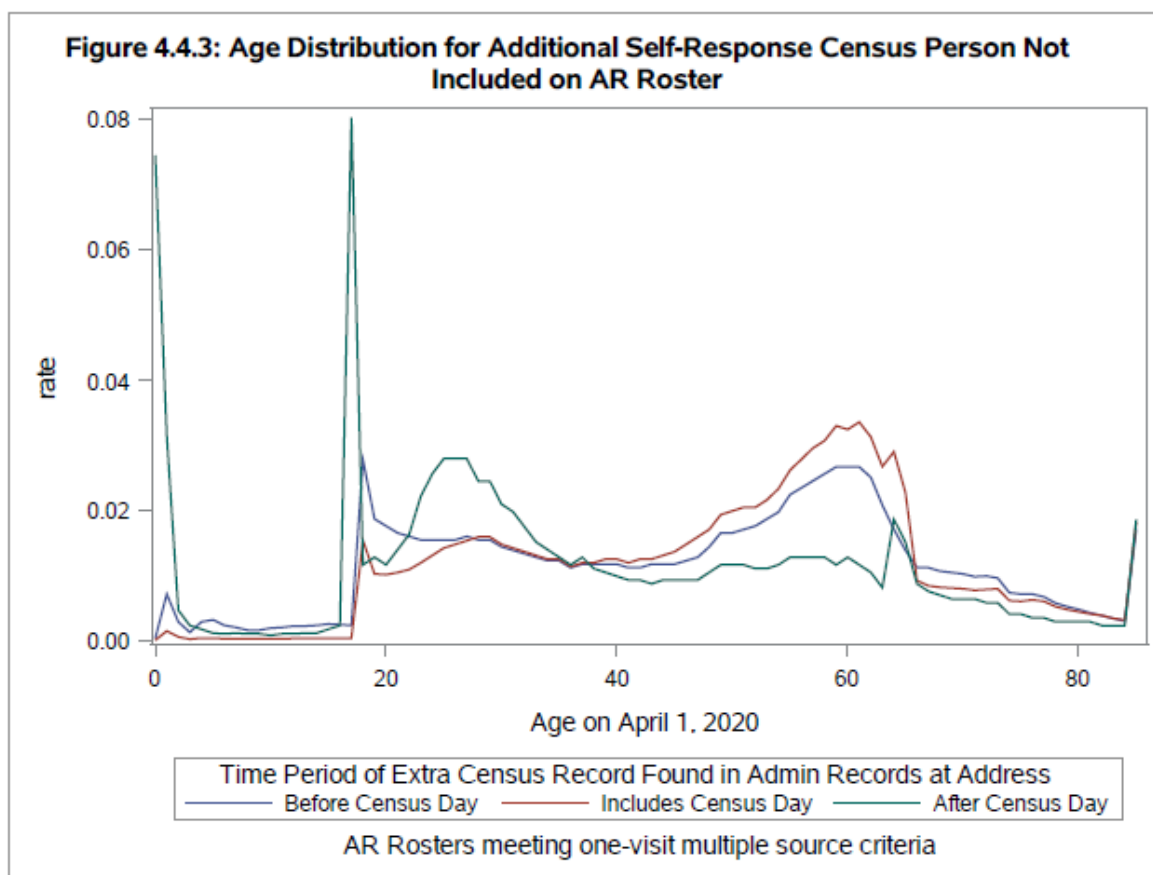
Note: Ages over 85 are truncated to 85 for disclosure avoidance.

Figure 4.4.3 shows two spikes that stand out in the age distribution for the persons on the census roster but not the AR roster who were found in the Self-response Verification file at the address after Census Day. Section 5 contains a discussion of several family situations that may be relevant to these spikes. Of the extra people found at the address only after Census Day, the proportion who are children ages 0 to 2 is 0.075. The proportion for the extra people found at the address only after Census Day who are ages 18 to 19 is about 0.08. The three lines have similar patterns for the remaining ages with two exceptions.

These two groups are children ages 0 to 2 and older children of age 17, both found at the address in the Self-response Verification file only after Census Day. The proportion for people of age 24 whose proportion appears to be about 0.03 of those found in the AR Composite at the address only after Census Day while the proportions in the other two categories for age 24 are both about 0.018. The other exception occurs about ages 55 to 64 that, when combined, appear to have a proportion of about 0.25 of the records that were found at the address on dates that include Census Day and a proportion of about 0.15 of the records were found at the address before Census Day.

Figure 4.4.4 contains two bar charts, one that adds the age distribution of the person missing on the AR roster for addresses with a NRFU household member response to those with a self-

response shown in Figure 4.4.3. The number of NRFU household member responses is small enough to require bar charts where the ages can be aggregated to meet the requirements for disclosure avoidance. Therefore, the ages for the person missing on the AR roster for self-responses also are aggregated to facilitate a comparison with the ages of the person missing on NRFU household member responses. The age distributions are shown in Figure 4.4.4 by whether records for the person were found in the Self-response Verification file before, after, or including Census Day. This arrangement enables a visual comparison of the patterns in the age distributions of the person missing from the AR roster for Self-response rosters and the NRFU household member response rosters. To the eyeball, there is a great deal of similarity in the distribution of the ages of those found on the census roster but not on the AR roster for the two types of responses under study. In addition, the age distributions for the two types of responses appear similar when the extra person was found at the address in the Self-response Verification file in the time periods Before Census Day and Including Census Day. Furthermore, when the extra person was found in the Self-response Verification file in the time period After Census Day, the age distributions for both types of responses appear similar to each other and different from the age distributions observed in the other two time periods. In particular, the percentage of the people over age 50 on the census roster but not the AR roster appears lower and the percentage of people ages 0 to 17 on the census roster but not the AR roster appears to be higher. However, no variances have been calculated for the percentages so statistical differences are not being inferred.

Figure 4.4.4: Age Distribution for Additonal Census Person Not Included on AR Roster
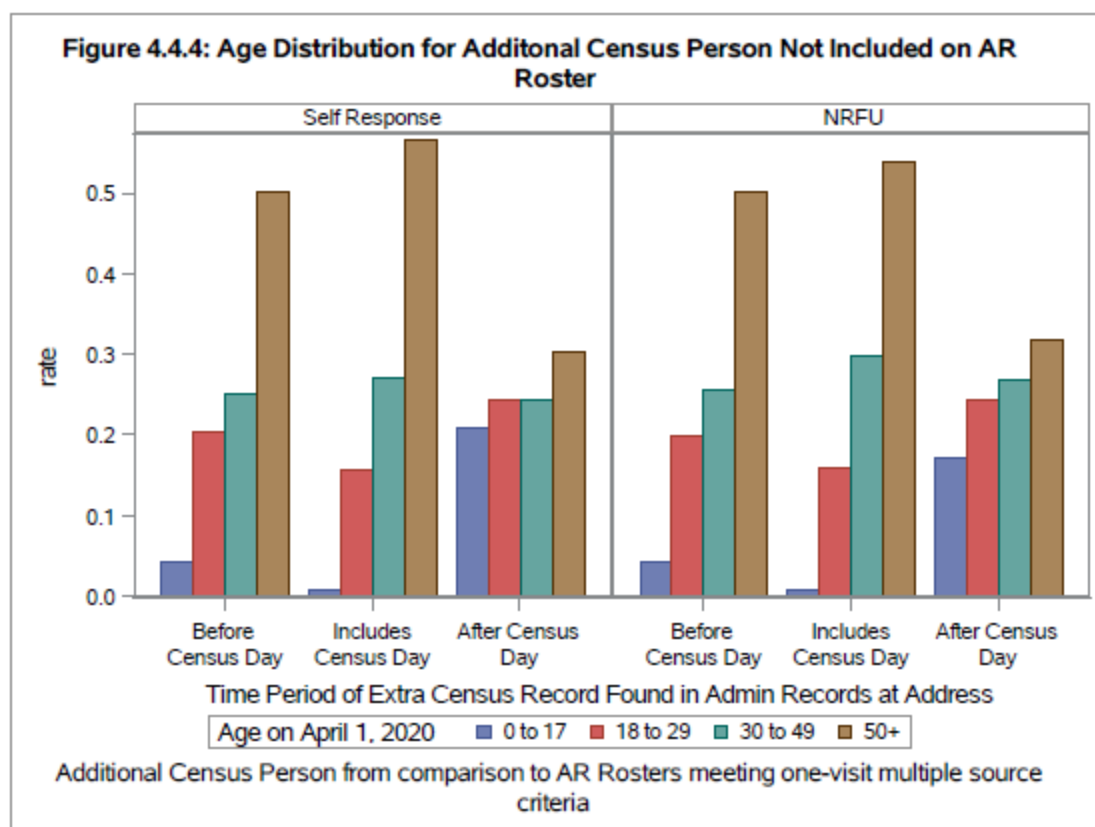
4.4.3 Relationship to householder when AR roster has one less than the census roster
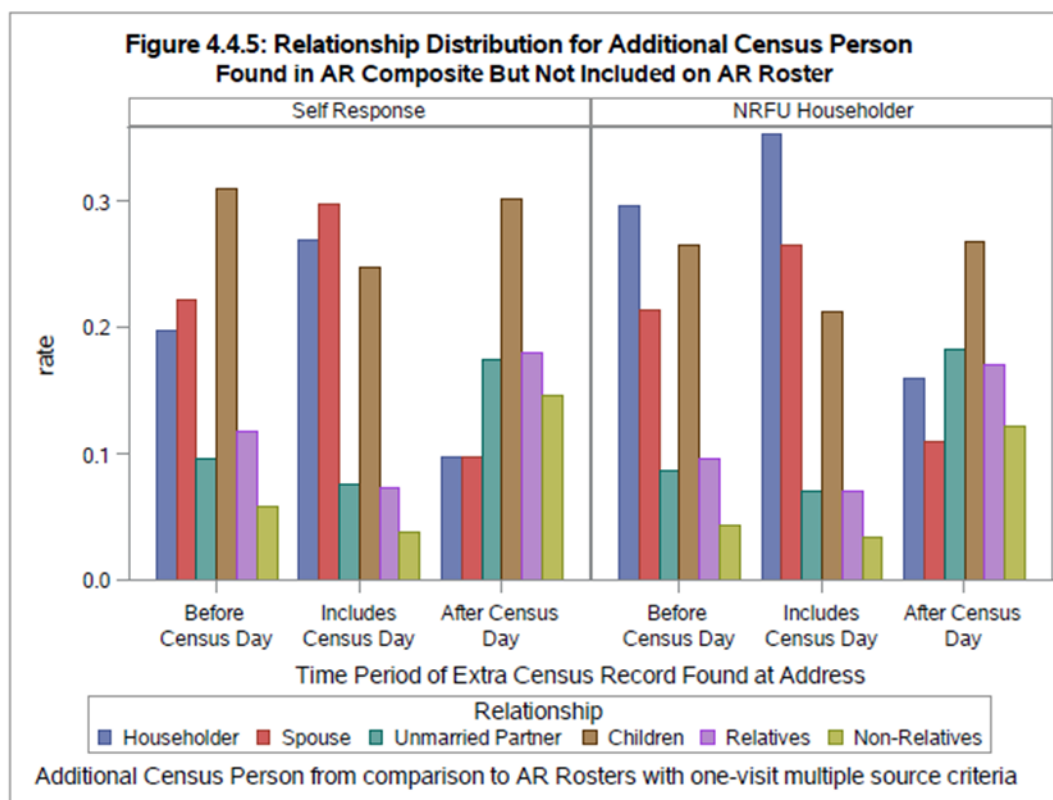
The analyses in this section focus on addresses where the AR roster has one less person than the census roster. The goal is to investigate the characteristics of the people on the census roster but not the AR roster by examining the relationship between the person and the householder. The census questionnaire includes a question that asks each person's relationship to the householder, which makes the analyses possible. However, the creation of the AR rosters relies mainly on the IRS 1040 and 1099 forms. These IRS forms do not ask about the relationship of dependents to the householder, which means the required data are not available for an analysis of the situations where the AR roster has one more person than the census roster. All the addresses included in the analyses have a One-Visit Multiple Source AR roster. This analysis uses data with final edit and imputed values.

Figure 4.4.5 shows the distributions of the relationship of the person not on the AR roster to the householder for self-responses and for NRFU household member responses by the date range of the AR(s) found for the person at the address in the Self-response Verification file. There are three date ranges, Before Census Day, Includes (before and after) Census Day, and After Census Day. The proportions for the types of relationship are calculated within each date range, which is indicated by the color of the column.

For the self-responses on the left in Figure 4.4.5, the proportion of the extra people on the census rosters who are the Householder at the address is 0.197 among those with an AR record found at the address Before Census Day in the Self-response Verification file. However, some results stand out in the date range Before Census Day. The proportion for the relationship Children is 0.310 and the proportion for Spouse is 0.221. The remaining three relationship groups in the Before Census Day group and their proportions are Unmarried Partners at 0.096, Relatives at 0.117, and Non-relatives at 0.058.

The distribution of the relationship to Householder within the date range Includes Census Day for self-responses has three relationship groups with noticeable proportions. These three groups and their proportions are Spouse at 0.298, Householder at 0.269, and Children at 0.248. The proportions for the other three groups are Unmarried Partner at 0.0798, Relatives at 0.073, and Non-relatives at 0.038.

The sizes of the groups in the date range After Census Day for self-responses appear to have a distribution that differs from the other two groups. The most prominent group is Children with a proportion of 0.302. The groups Relatives, Unmarried Partner, and Non-relatives have proportions of 0.180, 0.174, and 0.145, respectively. The remaining groups and their proportions are Householder at 0.098 and Spouse at 0.097.

Figure 4.4.5: Relationship Distribution for Additional Census Person Found in AR Composite But Not Included on AR Roster

Note: The rates are calculated by date group within response type.

Turning to the NRFU household member responses on the right in Figure 4.4.5, three relationship categories stand out in the date groups Before Census Day and Includes Census Day. In the date group Includes Census Day, the proportion for Spouse is 0.297 while the proportions for Householder and Children are 0.269 and 0.247, respectively. The same three categories and their proportions in the Before Census Day date group are Householder at 0.297, Children at 0.265, and Spouse at 0.214.

For the date group After Census Day among the NRFU household member responses, the relationship category Children is notable with a proportion of 0.304. Children has similar proportions for the date groups Before Census Day and Includes Census Day. However, for the date group After Census Day, the relationship groups Unmarried Partner and Relatives have proportions 0.179 and 0.176, respectively. The proportion for the Non-relatives category in the After Census Day date group is 0.145, while the proportions for Non-relatives in Before Census Day group and Includes Census Day group are 0.057 and 0.037, respectively.

In the Includes Census Day date group for the NRFU household member responses, the proportion for Householder is 0.352, The proportion for Spouse is 0.262, and the proportion for Children is 0.211.

For the After Census Day date group among the NRFU household member responses, the proportion who are Children is 0.265 with the proportions for Unmarried Partner and Relatives at

0.178 and 0.169, respectively. Interestingly, Children, Unmarried Partner and Relatives are also prominent Relationship groups in the After Census Day date group among the self-responses.

The similarity in patterns in the distributions of Relationships categories for One-Visit Multiple Source AR rosters for addresses with a census self-response or NRFU household members responses by date group may or may not have been expected. The self-responses and NRFU household member responses are thought to be the best census responses so possibly the similarities should be expected because of the high quality of both types of respondents.

4.4.4 Further analyses of addresses with householders on census roster but not AR roster

The next step is to investigate whether there is an association between the size of the household and the time period relative to Census Day when the householder is found at the address on a record in the Self-response Verification file. Figure 4.4.6 shows the distribution of the household size for the three time periods when a record was found at the address in the Self-response Verification file. Notice that for all three time periods, the two-person household size appears the most likely when the person on the census roster but not the AR roster is the householder since the proportions of the households having two people for the three time periods ranging from 0.68 to 0.82. The proportions for the three-person household size range from 0.13 to 0.18 while the proportions observed for the other households, which have four, five, and six or more people, range from about 0.01 to .09.

Figure 4.4.6: Household Size When Additional Census Person Not Included on AR Roster by Timeframe
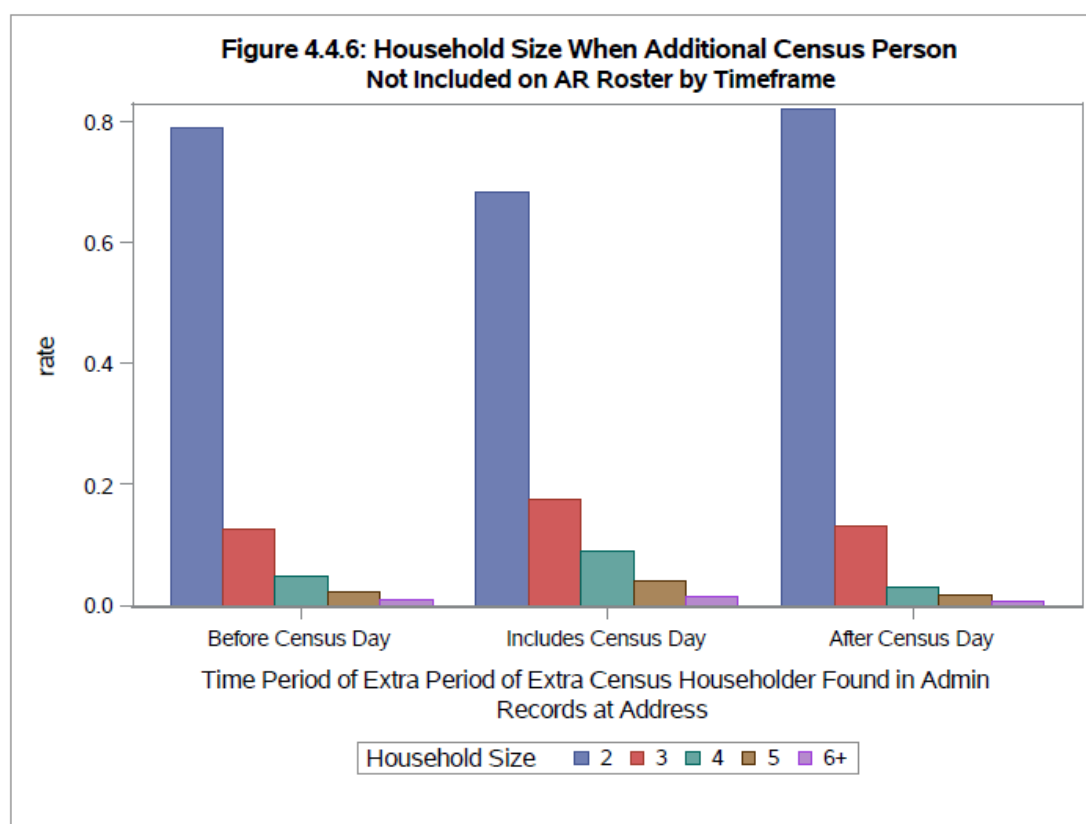
An additional analysis considers the age of the householder who is on the census roster but not the AR roster for the address and examines whether the distribution of the age of these householders appears to vary by the time period when an AR for the householder was found at the address. Figure 4.4.7 shows the age distribution for these householders by the time period when ARs for them were found in the Self-response Verification file. Notice that when ARs for the householder were found before Census Day or in the time period including Census Day, the proportion that were age 50 or older was about 0.71. However, for the householders with ARs found after Census Day, the proportion who are age 50 or older was about 0.35, which is about half of the proportions observed for the other two time periods.

Figure 4.4.7: Age Distribution of Extra Householder Census Person Not Included on AR Roster by Timeframe
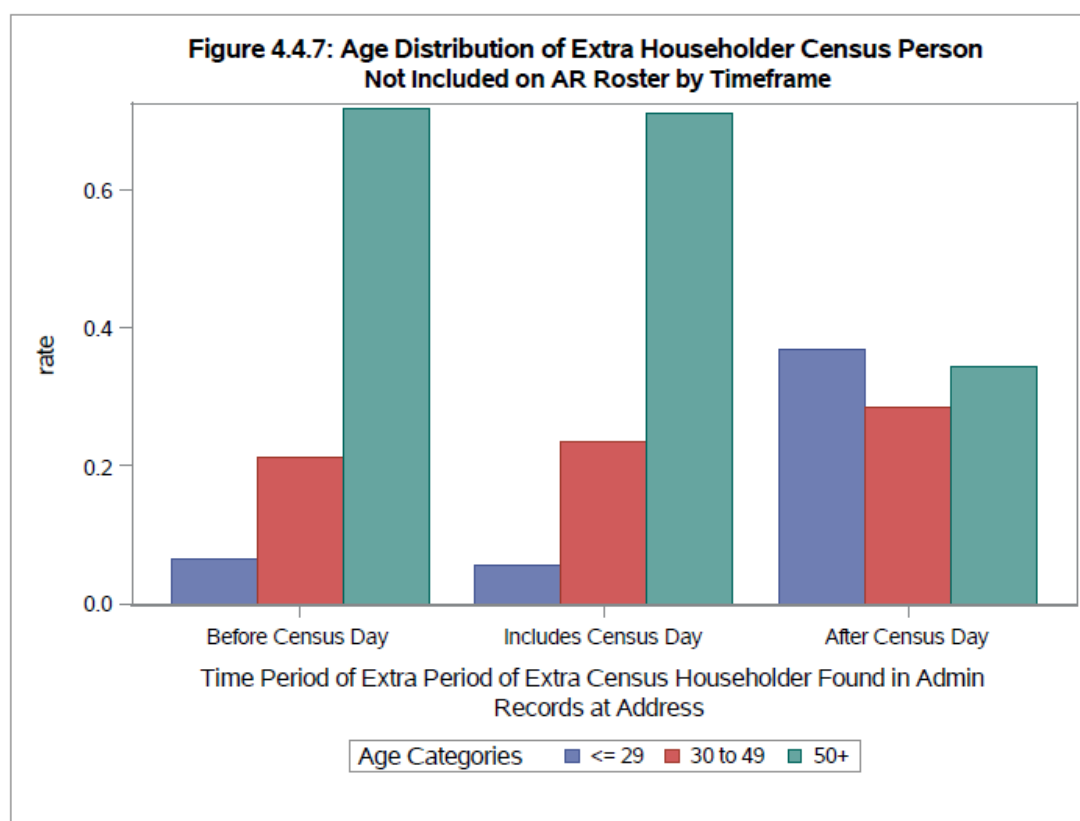
Even more interesting is that the proportions in Figure 4.4.7 for the householders who are either less than or equal to age 29 and those ages 30 to 49 are very similar for the time periods Before Census Day and Including Census Day but differ for the time period After Census Day. Both the Before and Including Census Day time periods have the proportion of about 0.06 for ages less than or equal to 29 and about 0.22 for ages 30 to 49. However, when the ARs were found After Census Day, the proportion that are householders of age 29 or less is 0.37 while the proportion for householders of age 30 to 49 is 0.29.

More research is needed to explain the difference in the distribution of the age of the householders found at the address in Self-response Verification file After Census Day from those found in Self-response Verification file Before or Including Census Day. However, the COVID-19 pandemic caused many people to move around Census Day because of such things as

universities closing their dormitories, evictions because of people losing their jobs and not being able to pay rent, and an exodus from the large cities to less populated areas to avoid exposure to the virus.

The last analysis focuses on the two-person households where the householder is included on the census roster but not the AR roster for the address. In particular, the analysis examines the distribution of the relationship of the second person in the household to the householder and whether that relationship varies by the time period when an AR record for the householder was found at the address in the Self-response Verification file.

Figure 4.4.8 displays the distribution of the relationship of the second person to the householder who is on the census roster but not the AR roster for the three time periods when a record(s) for the householder was found at the address in the Self-response Verification file. The distributions for the second person's relationship to householder appear similar when an AR record(s) for the householder was found at the address in the date range Before Census Day or Including Census Day. When the AR for the householder was found Before Census Day, the proportion of the second people who are a spouse is about 0.54 while the same proportion is 0.58 when the AR for the householder was found in the date range Including Census Day. The other relationships for the second person appear to occur at similar rates when the AR record for the householder was found before Census Day or Including Census Day.

Figure 4.4.8: Distribution of Second Person Relationship Type for 2 Person Households Where Extra Person was Householder
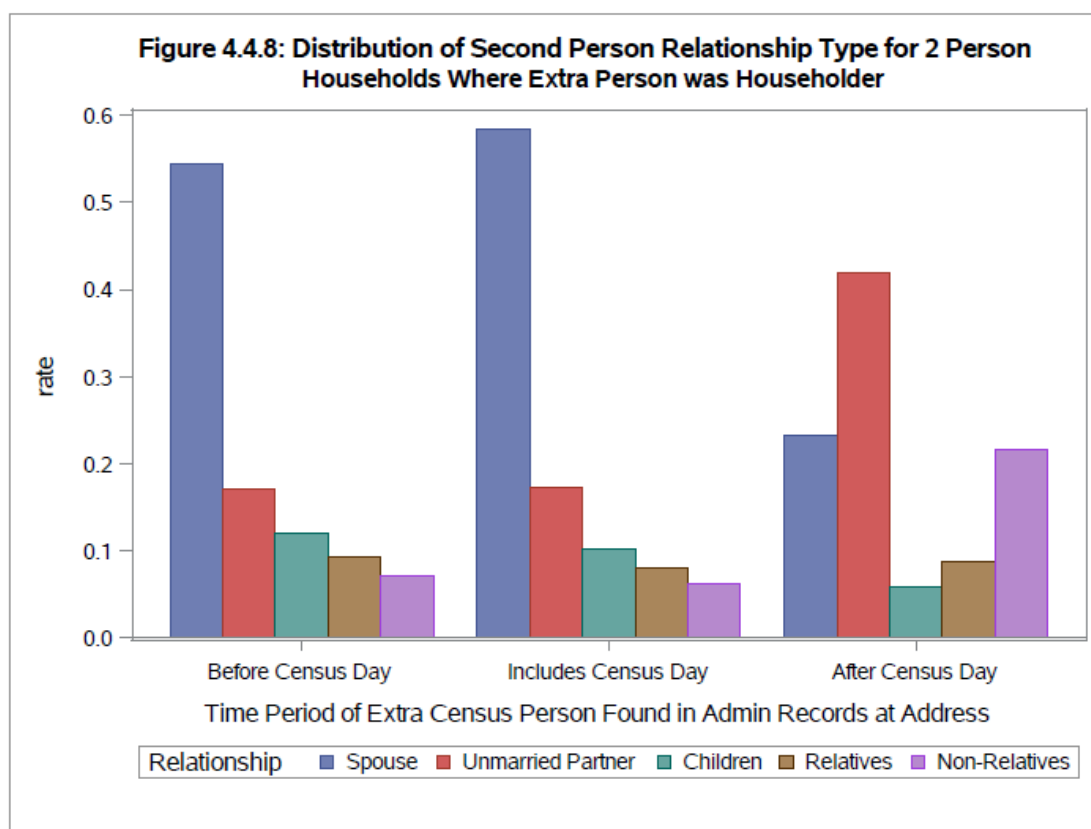
However, if a record for the householder was found the Self-response Verification file After Census Day, the relationship of the second person appears more likely to be an unmarried partner

and much less likely to be a spouse. For example, the proportion of the second persons classified as an Unmarried Partner of the householder is about 0.42 while the proportion of second persons with the relationship of spouse is about 0.23. In addition, the proportion of second persons with the relationship to the householder of Non-relative is about 0.22. The proportion classified as Children is about 0.06 while the proportion classified as Relatives is 0.09

5. Discussion

The focus of this section is on some plausible explanations for why the census roster and the AR roster for an address differ by one person for some addresses. Examining some of the IRS rules about claiming dependents offers some insight as to how a person may be on one of the rosters but not the other. Another avenue to consider is the effects of the COVID-19 pandemic causing delays in census data collection and the delivery of IRS data to the Census Bureau, which resulted in a delay of the AR processing for the census. The intent of this section is not to give an extensive review of IRS rules for claiming dependents, but to identify some of the rules that may help explain why the census roster and the AR roster differ by one. This section also examines some of the reasons related to 2020 Census processing of the administrative records that could lead to these differences as well.

5.1 IRS rules for claiming dependents affecting children and relatives

The IRS has rules about whom a taxpayer can claim as a dependent on a 1040 tax return. There are two basic sets of rules for when a taxpayer may claim a person as a dependent, one set applies to children and the other applies to relatives. There are situations when a child does not qualify as a dependent child but does qualify as a dependent relative. Usually, a dependent must reside with the household of the taxpayer all year, but there are situations where a close relative does not have to reside with the household. However, the taxpayer does have to provide support for the dependent for at least half of the year.

5.1.1 Claiming a child or other relative as a dependent

A taxpayer may claim his/her child as a dependent if the child meets either the Qualifying Child Test or the Qualifying Relative Test. Sometimes a taxpayer's child may not meet the requirements to be a Qualifying Child but does meet the requirements to be a Qualifying Relative. A Qualifying Child and Qualifying Relative must be a U.S. citizen, U.S. resident alien, U.S. national, or a resident of Canada or Mexico in addition to meeting the criteria listed below:

- To meet the Qualifying Child Test, the child must be younger than the taxpayer and either younger than 19 years old or be a student younger than 24 years old as of the end of the calendar year. However, there is no age limit if the child is permanently and totally disabled.
    - The child must be the taxpayer's son, daughter, stepchild, foster child, brother, sister, half-brother, half-sister, stepbrother, stepsister, or a descendant of any of these.

- To meet the Qualifying Relative Test, the relative must meet one of the following requirements
  - The person cannot be taxpayer's qualifying child or the qualifying child of any other taxpayer.
  - The person either (a) must be related to taxpayer in one of the ways listed in Table 5.1, or (b) must be a member of the taxpayer's household all year (and the relationship must not violate local law).
  - The person's gross income for the year must be less than $4,300.
  - The taxpayer must provide more than half of the person's total support for the year.

Table 5.1. Children and Other Relatives who do not have to be a member of the taxpayer's household all year to be claimed as a dependent.

The taxpayer's child, legally adopted child, stepchild, foster child, or a descendant of any of them (for example, the taxpayer's grandchild).
The taxpayer's brother, sister, half-brother, half-sister, stepbrother, or stepsister.
The taxpayer's father, mother, grandparent, or other direct ancestor, but not foster parent.
The taxpayer's stepfather or stepmother.
A son or daughter of the taxpayer's brother or sister.
A son or daughter of the taxpayer's half-brother or half-sister.
A brother or sister of the taxpayer's father or mother.
The taxpayer's son-in-law, daughter-in-law, father-in-law, mother-in-law, brother-in-law, or sister-in-law.

Note: See Internal Revenue Service (2020)

Any of the relationships listed in Table 5.1 that were established by marriage are not ended by death or divorce. For example, if a taxpayer's spouse dies and the taxpayer is supporting the spouse's father, the taxpayer still can claim the father-in-law as a dependent if the father-in-law meets the other criteria.

A person is considered to live with the taxpayer as a member of the taxpayer's household during periods of time when the person, the taxpayer, or both are temporarily absent because of special circumstances, such as illness, education, business, vacation, military service, or detention in a juvenile facility. In addition, if the person is placed in a nursing home for an indefinite period of time to receive constant medical care, the absence may be considered temporary.

5.1.2 Reasons for Children and Other Relatives not on both rosters

The rule that a taxpayer can claim the taxpayer's child as a dependent if the child is age 18 or younger at the end of the tax year and not disabled explains the spike in Table 4.4.3 for children ages 19 to 24 who are not students not being listed as dependent on the parent's IRS 1040. However, these individuals may appear on a separate IRS 1040 or 1099 at the parent's address. Even if the child is 19 or older and not a student, the child may use the parents' address to receive some types of mail when not living in the parent's household.

Some of the people over age 24 who do not appear on an IRS 1040 may live at the address but do not qualify to be claimed as dependent. In addition, the person may have received a 1099, but their income is too low to be required to file an IRS 1040 form. Or, a person, may not live at the address but uses the address for mail and received an IRS 1099 form. An example is a person living in a nursing home or other care center.

In addition, a person of age 18 or younger with divorced parents may be claimed as a dependent at the non-custodial parent's address if the custodial parent agrees as part of the divorce proceedings. However, the child may receive some mail, such as for a savings account, at the custodial parent's address. This may explain why some children of age 18 or younger were not listed on the IRS 1040 for the custodial parent's address but received am IRS 1099 at the address or were in other ARs at the custodial parent's address. Or, if the custodial parent claims the child as a dependent, the child could receive an IRS 1099 for a savings account or another type of account, at the non-custodial parent's address, particularly if the non-custodial parent is the one who set up the account.

Similar arguments apply for Other Relatives who live at the taxpayer's address but do not qualify as the taxpayer's dependent. For example, the taxpayer may pay less than half of the Other Relative's total support for the year. Then, the Other Relative would be listed on the census roster but probably not on the AR roster.

5.2 Linking Young Children to Parents

Figure 4.4.1 shows a spike in the number of children ages 0 to 4 who were not listed on AR rosters. The bulk of these appear to be newborns in 2020. Children born on or before April 1, 2020, were eligible for census enumeration. Children born after Census Day, which was April 1, 2020, were not eligible for enumeration since the goal of the census is to count resident population on April 1 of the census year.

The Social Security Administration delivers the Numident file to the Census Bureau on a quarterly basis. The Numident files contain new applications for Social Security numbers (SSNs). These applications are for newborns and immigrants to the U.S. who qualify for being assigned a SSN. The Census Bureau uses this information about the child, mother, and father to create the Household Composition Key file that can link children to their parents. The original processing plan for producing the response records for AR enumerations did not include births in 2020 before Census Day. This information was delivered to the Census Bureau in 2020 Quarter 2 Numident delivery. One of the lessons learned from the 2020 Census processing is that

additional processing can be added to insert these children into their parent's rosters. This would entail a) creating an update to the Household Composition Key File to enable linking these children to their parents and b) adding children to the AR rosters at addresses where the parents lived when appropriate. However, if the AR roster for an address already has six people, adding the child will result in in the AR roster having seven people and disqualify it for AR enumeration.

5.3 Householder and Spouse not on AR rosters

Householders and Spouses have higher percentages of those on the census roster but not the AR roster than one might expect as shown in Figure 4.4.4 for the time periods Before Census Day and Includes Census Day. The percentages for Householders and Spouses that are on the census roster but not the AR roster appear lower for the After Census Day time period.

One explanation is that the Householder and Spouse may have filed separately. Possibly one of the IRS 1040s arrived while the AR processing was incorporating new IRS data and the other was submitted too late to be included in the AR roster for the address.

5.4 Unmarried Partners and Non-relatives not on AR rosters

Unmarried Partners and Non-relatives have the reverse pattern to the one observed for Householders and Spouses. The Unmarried Partners and Non-relatives have higher percentages of those on the census roster but not the AR roster in the After Census Day time period than in the Before Census Day and Includes Census Day time periods.

5.5 Impact on census enumeration

The source of the disagreement between the census rosters and AR rosters on household size for some addresses appears to be related a couple of factors. One factor is that some members of households filed 1040 tax returns in October 2020 or later, which meant the AR modeling process stopped before those returns could be delivered. Another factor appears to be the IRS rules about whom a taxpayer can claim as a dependent. Although a taxpayer is required to pay for at least half of a relative's support to claim the relative as a dependent, there are situations where the relative does not have to reside in the person's household at least half of the year. Another situation involves children in joint custody where the divorce agreement specifies that the non-custodial parent may claim the child as a dependent although the child resides with the custodial parent more than half the year. In these two situations, the taxpayer would claim the relative, including children, as a dependent at the taxpayer's address, but the relative's Census Day address would be location of the enumeration, which is where the person resides more than half the year.

6. Summary

The evaluation project focused on providing information about AR modeling and AR enumeration that will aid in planning the 2030 Census. In particular, the research concentrated on addresses that had both an AR roster and a census roster and examined whether the two rosters

agreed on household size. The analysis was restricted to census self-responses and Nonresponse Followup (NRFU) household member responses because these are the highest quality of census responses. The census rosters were partitioned by the timing of the submission of the response, which were before July 30 and after July 30 for self-responses and after July 30 for NRFU household member responses. Also considered was the quality category of the AR rosters available for enumerating the address, which are One-Visit Multiple Source, Closeout Multiple Source, and Closeout Household Size Only, listed in the order of their quality,

A key result from Table 4.1.5 showed that for addresses with both a One-Visit Multiple Source AR roster and a census self-response roster submitted by July 30, the two rosters had a household size agreement rate of 83.8 percent. This category has 44.54 million addresses. For the 1.957 million addresses with the same type of AR roster and a census self-response dated after July 30, the household size agreement rate for the two rosters is 74.4 percent. The category that has addresses with the same type of AR roster and a NFRU household member response has 2.083 million addresses and a household size agreement rate of 66.3 percent. The average household size for the One-Visit Multiple Source AR rosters is 2.2 people.

Also in Table 4.1.5, the household size agreement rate observed for 4,431,000 addresses with a Closeout Multiple Source AR roster and a self-response census roster submitted before July 30 is 51.8 percent. The agreement rate when the self-response census roster was submitted after July 30 was 51.6 percent. while the agreement rate when the census response was submitted during NRFU was 41.3 percent.

When turning to the results for the Household Size Only AR rosters, the addresses that have one of these rosters and a census self-response submitted before July 30 have a household size agreement rate of 68.2 percent. For addresses with a self-response submitted after July 30 for an address with the same type of AR roster, the agreement rate was 64.6 percent. For addresses with a NRFU household member respondent and a Household Size Only AR roster, the agreement rate for household size is 53.3 percent.

The combined results of the research indicate that both the mode of response and the amount recall required of the respondent because of the length of time since April 1 affect the agreement rate between the census roster and the AR roster. In addition, the results show the value ARs can bring to improvements in census enumeration. More research is needed to identify the important factors in creating AR rosters for census enumeration and to refine the methods used in constructing the AR rosters.

Section 4.2 provided more information about what types of census household rosters had a higher count than the administrative record rosters. The results showed that households with other relatives or non-relatives present stood out. One possible modification of the AR modeling methodology for the 2020 Census is to include an additional model that would be used if the census roster has other relatives or non-relatives present given the AR people and characteristics observed for the address. Households with larger probabilities of having other relatives or non-relatives present may be candidates for more contacts.

For addresses where the census roster and the AR roster sizes differed by only one person, the additional analysis was able to highlight the age, sex and relationship characteristics of the person included on one of the rosters but not the other. The analysis was able to use historical person and address information to quantify how often the additional person was associated with the roster address or other addresses and the timing of these associations in relation to Census Day. This information has the potential of being useful in the research aimed at finding better ways to account for these types of differences as part of the 2030 Census planning and may include identifying other sources and rostering methods to use in the future.

7. References

Census Scientific Advisory Committee (2020). Video of Meeting on September 30, 2020. Washington, DC: U.S. Census Bureau.

Deaver, K. D. (2020). *Intended Administrative Data Use in the 2020 Census*. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-data-use-2020-census.pdf

Internal Revenue Service (2020). *Dependents*. Washington, DC: Internal Revenue Service. https://www.irs.gov/faqs/filing-requirements-status-dependents/dependents

Keller, A. (2019). "Analyzing Tradeoff Between Administrative Records Enumeration and Count Imputation." *Proceedings of the 2019 Joint Statistical Meeting, Government Statistics Section*. Alexandria, VA: American Statistical Association.

Keller, A., T. Mule, D. S. Morris, and S. Konicki (2018). A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census. *Journal of Official Statistics*, Vol. 34, No. 3, 2018, pp. 599–624, http://dx.doi.org/10.2478/JOS-2018-0029

Morris, D. S. (2017), "A Modeling Approach for Administrative Record Enumeration in the Decennial Census," *Public Opinion Quarterly*, 81(S1), 357–384. https://doi.org/10.1093/poq/nfw059 Accessed January 2023.

Morris, D. S. (2014). "A Comparison of Methodologies for Classification of Administrative Records Quality for Census Enumeration." *Proceedings of the 2014 Joint Statistical Meeting, Survey Research Methods Section*. Alexandria, VA: American Statistical Association. 1729-1743. http://www.asasrms.org/Proceedings/y2014/files/311864_88281.pdf

Mule, T., A. K. Keller, and S. Konicki (2018). "Administrative Record Modeling Research Synthesis." DSSD 2020 DECENNIAL CENSUS R&T MEMORANDUM SERIES #L-15. Washington, DC: U.S. Census Bureau.

Mulrow, E., A. Mushtaq, S. Pramanik, and A. Fontes (2011). *Assessment of the U.S. Census Bureau's Person Identification Validation System*. Chicago, IL: NORC.

https://www.norc.org/PDFs/May%202011%20Personal%20Validation%20and%20Entity%20Resolution%20Conference/PVS%20Assessment%20Report%20FINAL%20JULY%202011.pdf

Mulry, M. H., T. Mule, A. K. Keller, and S. Konicki (2021). "Overview of Administrative Records Modeling in the 2020 Census**."** 2020 CENSUS PROGRAM MEMORANDUM SERIES: 2021.10. Washington, DC: Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf

Rastogi, S. and A. O'Hara (2012). "2010 Census Match Study." 2010 Census Planning Memorandum No. 247. Washington, DC: U.S. Census Bureau. https://www.census.gov/library/publications/2012/dec/2010_cpex_247.html

Wagner, D. and M. Layne (2014). "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications." CARRA Working Paper Series. Working Paper #2014-01. Washington, DC: Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf

U.S. Census Bureau (2021). 2020 Census – Release 3_Table 2.xlsx. In 2020 Census Data Quality Operational Metrics: Release 3.xlsx. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/data/operational-quality-metrics/census-operational-quality-metrics-release_3-table2.xlsx

U. S. Census Bureau (2019). "2020 Census Detailed Operational Plan for: 19. Response Processing Operation (RPO)." Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/RPO_detailed_operational_plan-v2.pdf

U.S. Census Bureau Administrative Records Modeling Team (2017). *Administrative Records Modeling Update for the Census Scientific Advisory Committee.* Washington, DC: U.S. Census Bureau.
https://www2.census.gov/cac/sac/meetings/2017-03/admin-records-modeling.pdf

Appendix. Table A1. Percentage of addresses with census roster larger than AR roster by census roster trait when census household size differs from the household size for three types of AR rosters discussed in Section 4.2.

Census Roster Trait	One-visit multiple source			Closeout multiple source			Closeout Household Size Only		
	Total	Column Percent	Percent Census is higher when roster sizes are different	Total	Column Percent	Percent Census is higher when roster sizes are different	Total	Column Percent	Percent Census is higher when roster sizes are different
Overall	8,397,000	100.0%		2,687,000	100.0%		1,813,000	100.0%	
NH White Alone Householder	5,680,000	67.6%	55.2%	1,712,000	63.7%	24.6%	836,000	46.1%	77.0%
NH Black Alone Householder	911,000	10.8%	60.9%	348,000	13.0%	34.4%	302,000	16.7%	71.0%
NH AIAN Alone Householder	32,500	0.4%	61.9%	13,000	0.5%	41.2%	10,500	0.6%	75.4%
NH Asian Alone Householder	432,000	5.1%	62.1%	131,000	4.9%	43.1%	139,000	7.7%	81.1%
NH NHPI Alone Householder	7,400	0.1%	64.1%	2,800	0.1%	47.0%	3,700	0.2%	74.2%
NH SOR Alone Householder	35,500	0.4%	59.6%	11,500	0.4%	38.2%	10,500	0.6%	77.4%
NH 2 or More Races Householder	299,000	3.6%	60.7%	94,000	3.5%	33.5%	77,500	4.3%	77.1%
Hispanic Householder	1,000,000	11.9%	66.4%	374,000	13.9%	46.8%	433,000	23.9%	79.1%
At least One NH White Alone Person	6,027,000	71.8%	56.7%	1,810,000	67.4%	25.9%	936,000	51.6%	78.7%
At least One NH Black Alone Person	1,036,000	12.3%	64.0%	382,000	14.2%	36.5%	344,000	19.0%	73.7%
At least One NH AIAN Alone Person	57,500	0.7%	71.9%	20,500	0.8%	46.7%	17,000	0.9%	81.2%
At least One NH Asian Alone Person	551,000	6.6%	66.4%	160,000	6.0%	44.3%	166,000	9.2%	83.8%
At least One NH NHPI Alone Person	14,500	0.2%	73.6%	4,700	0.2%	51.7%	5,600	0.3%	78.9%
At least One NH SOR Alone Person	62,500	0.7%	70.6%	19,500	0.7%	47.3%	19,000	1.0%	82.8%
At least One NH 2 or More Races Person	599,000	7.1%	69.3%	183,000	6.8%	41.8%	144,000	7.9%	82.0%
At least One Hispanic	1,347,000	16.0%	69.9%	474,000	17.6%	47.5%	523,000	28.8%	80.9%
Spouse Present	3,605,000	42.9%	73.3%	1,236,000	46.0%	39.7%	580,000	32.0%	89.5%
Unmarried Partner Present	785,000	9.3%	87.6%	191,000	7.1%	50.6%	341,000	18.8%	92.8%
Children Present	3,642,000	43.4%	75.1%	1,140,000	42.4%	54.1%	803,000	44.3%	83.9%
Other Relatives Present including Grandchildren	1,247,000	14.9%	93.5%	314,000	11.7%	74.2%	302,000	16.7%	94.3%
Non-Relatives Present	710,000	8.5%	94.8%	144,000	5.4%	73.9%	311,000	17.2%	96.0%
At least one person age 0 to 4	967,000	11.5%	82.2%	306,000	11.4%	61.5%	333,000	18.4%	84.4%
At least one person age 5 to 17	2,008,000	23.9%	64.4%	799,000	29.7%	52.9%	516,000	28.5%	82.2%
At least one person age 18 to 24	1,346,000	16.0%	85.2%	390,000	14.5%	64.7%	534,000	29.5%	88.5%
At least one person age 25 to 49	4,511,000	53.7%	69.2%	1,418,000	52.8%	44.2%	1,314,000	72.5%	79.0%
At least one person age 50 to 64	3,769,000	44.9%	62.0%	1,108,000	41.2%	31.2%	499,000	27.5%	80.2%
At least one person age 65+	2,616,000	31.2%	61.6%	736,000	27.4%	25.1%	200,000	11.0%	83.6%

Note: Source of characteristics is the census roster.